Université Pierre et Marie Curie
Licence de Mathématiques ($3^{\text{ème}}$ année)
Année 2004/2005

Probabilités
Pierre Priouret

Mode d'emploi

Ce polycopié est destiné aux étudiants de la Licence ($3^{\text{ème}}$ année) de Mathématiques de l'Université Pierre et Marie Curie. En principe ces étudiants ont déja suivi un cours de théorie de la mesure et d'intégration. Nous commençons par l'étude des probabilités sur les ensembles finis (chapitre 1) puis sur les ensembles dénombrables (chapitre 2) avant de présenter (chapitre 3) les résultats d'intégration utilisés par la suite. Le chapitre 4 introduit les principales notions de probabilités dans leur cadre général. Le chapitre 5 traite des fonctions caractéristiques et des vecteurs gaussiens. Les théorèmes limites sont abordés dans les chapitres 6 (avec, en particulier, la loi des grands nombres) et 7 (avec, en particulier, la convergence en loi vers la loi normale). Enfin le chapitre 8 présente quelques notions de statistique.

Les compléments situés à la fin de certains chapitres ne sont pas au programme de l'examen.

Ce polycopié est divisé en chapitres, sections et sous-sections. Ainsi 3.2.4 renvoie au chapitre 3, section 2, sous-section 4 et 5.4 renvoie chapitre 5, section 4. A l'intérieur d'une même section, les énoncés sont numérotés en continu. Ainsi "d'après le th. 5.4.6" renvoie au chapitre 5, section 4, énoncé 6. Quant aux égalités, elles sont numérotées entre parenthèses et en continu au sein d'un même chapitre. Ainsi "vu (3.5)" réfère à la cinquième égalité numérotée du chapitre 3. Le signe $\diamond$ indique la fin d'une preuve. Ce polycopié se termine par un index des notations et un index des termes.

Table des matières

Chapitre 1

Espace de probabilité fini

Dans ce premier chapitre, on présente les premières notions de probabilité dans un cadre élémentaire.

1.1. Notions fondamentales

1.1.1. Probabilité sur un ensemble fini. Soit E un ensemble fini. Une probabilité sur E est une famille $(p(a),\, a \in E)$ de réels vérifiant

$$0 \leq p(a) \leq 1, \quad \sum_{a \in E} p(a) = 1.$$

On pose alors, pour $A \subset E$, $\mathbb{P}(A) = \sum_{a \in A} p(a)$. $\mathbb{P}$ est une application de $\mathcal{P}(E)$ dans $[0,1]$ telle que

$$\mathbb{P}(\Omega) = 1, \;\; \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) \text{ si } A \cap B = \emptyset. \tag{1.1}$$

On voit immédiatement, par récurrence, que, si $A_1, \ldots, A_r$ sont des sous-ensembles de Ω deux à deux disjoints, alors

$$\mathbb{P}(\bigcup_{i=1}^{r} A_i) = \sum_{i=1}^{r} \mathbb{P}(A_i).$$

Réciproquement si une fonction d'ensembles $A \mapsto \mathbb{P}(A)$, $A \subset E$, vérifie (1.1) et si on pose, pour tout $a \in E$, $p(a) = \mathbb{P}(\{a\})$, on a $0 \leq p(a) \leq 1$ et $\sum_{a \in E} p(a) = 1$ puisque les ensembles $\{a\}$ sont évidemment deux à deux disjoints d'union E. En conclusion, on appellera probabilité sur E aussi bien la famille $(p(a),\, a \in E)$ que la fonction d'ensembles $A \mapsto \mathbb{P}(A)$.

1.1.2. Espace de probabilité fini. Un couple $(\Omega, \mathbb{P})$ où Ω est un ensemble fini et $\mathbb{P}$ une probabilité sur Ω s'appelle un espace de probabilité fini. Un sous-ensemble A de Ω s'appelle un événement et $\mathbb{P}(A)$ est la probabilité que l'événement A ait lieu. L'élément $\{\omega\}$ s'appelle alors un événement élémentaire. On note A^c le complémentaire de A,

c'est l'événement "A n'a pas lieu". De même $A \cup B$ est l'événement "A ou B a lieu" et $A \cap B$ est l'événement "A et B ont lieu". Enfin Ω est l'événement certain et $\emptyset$ est l'événement impossible. Noter (c'est la moindre des choses) que $\mathbb{P}(\emptyset) = 0$ puisque, vu que $\Omega \cap \emptyset = \emptyset$,

$$1 = \mathbb{P}(\Omega) = \mathbb{P}(\Omega \cup \emptyset) = \mathbb{P}(\Omega) + \mathbb{P}(\emptyset) = 1 + \mathbb{P}(\emptyset).$$

Donnons quelques conséquences faciles de (1.1). On a $A \cup A^c = \Omega$ et $A \cap A^c = \emptyset$ donc $1 = \mathbb{P}(\Omega) = \mathbb{P}(A) + \mathbb{P}(A^c)$ d'où

$$\mathbb{P}(A^c) = 1 - \mathbb{P}(A). \tag{1.2}$$

Si $A \subset B$, on note $B \setminus A = B \cap A^c$. On a alors $B = A \cup (B \setminus A)$ avec $A \cap (B \setminus A) = \emptyset$ d'où

$$\text{si } A \subset B,\ \ \mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A). \tag{1.3}$$

En particulier, dans ce cas, $\mathbb{P}(A) \leq \mathbb{P}(B)$. Enfin on a

$$A \cup B = (A \cap B) \cup (A \setminus A \cap B) \cup (B \setminus A \cap B),$$

ces ensembles étant deux à deux disjoints. On a donc

$$\mathbb{P}(A\cup B) = \mathbb{P}(A\cap B)+\mathbb{P}(A\setminus A\cap B)+\mathbb{P}(B\setminus A\cap B) = \mathbb{P}(A\cap B)+\mathbb{P}(A)-\mathbb{P}(A\cap B)+\mathbb{P}(B)-\mathbb{P}(A\cap B)$$

d'où

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B). \tag{1.4}$$

On note $|A|$ le cardinal de A i.e. le nombre d'éléments de A. Un cas particulier important d'espace de probabilité fini $(\Omega, \mathbb{P})$ est celui où $\mathbb{P}$ est la probabilité uniforme sur Ω définie par

$$\mathbb{P}(\{\omega\}) = \frac{1}{|\Omega|}.$$

On a alors $\mathbb{P}(A) = \frac{|A|}{|\Omega|}$. Ce cas est très fréquent mais n'est pas le seul à envisager (voir l'exemple 4 de 1.1.4).

1.1.3. Variables aléatoires. Soit $(\Omega, \mathbb{P})$ un espace de probabilité fini. On appelle variable aléatoire (en abrégé v.a.) à valeurs E toute application X de Ω dans E. Puisque $X(\Omega)$ est fini, on peut supposer E fini, c'est ce qu'on fera par la suite. Pour $a \in E$ et $\Gamma \subset E$, on pose

$$\{X = a\} = X^{-1}(a) = \{\omega,\ X(\omega) = a\},\ \ \{X \in \Gamma\} = X^{-1}(\Gamma) = \{\omega,\ X(\omega) \in \Gamma\}. \tag{1.5}$$

On définit alors, pour tout $a \in E$, $q(a) = \mathbb{P}(X = a)$. On a $0 \leq q(a) \leq 1$ et, les ensembles $\{X = a\}$, $a \in E$, étant deux à deux disjoints d'union Ω, $\sum_{a\in E} q(a) = \mathbb{P}(\Omega) = 1$. Les $(q(a),\, a \in E)$ sont donc une probabilité sur E, notée μ_X, appelée loi de la v.a. X. Alors, pour tout $\Gamma \subset E$,

$$\mu_X(\Gamma) = \sum_{a\in\Gamma} q(a) = \sum_{\omega,\ X(\omega)\in\Gamma} p(\omega) = \mathbb{P}(X \in \Gamma).$$

1.1.4. Exemples.

1. On lance une pièce trois fois de suite. L'ensemble des issues possibles est

$$\Omega = \{PPP, PPF, PFP, PFF, FPP, FPF, FFP, FFF\}.$$

On a $|\Omega| = 2^3 = 8$. Les issues étant équiprobables, on munit Ω de la probabilité $\mathbb{P}(\{\omega\}) = \frac{1}{8}$. Soient A l'événement "on obtient exactement deux faces" et B l'événement "on obtient au moins deux faces". On a $A = \{PFF, FPF, FFP\}$, $B = \{PFF, FPF, FFP, FFF\}$, $|A| = 3$, $|B| = 4$, $\mathbb{P}(A) = \frac{3}{8}$, $\mathbb{P}(B) = \frac{1}{2}$.

2. On lance deux dés, un rouge et un bleu. L'ensemble des issues possibles est

$$\Omega = \{11, 21, 12, \ldots, 66\} = \{i_1 i_2,\ 1 \leq i_1, i_2 \leq 6\}.$$

On a $|\Omega| = 6^2 = 36$. Les issues étant équiprobables, on munit Ω de la probabilité $\mathbb{P}(\{\omega\}) = \frac{1}{36}$. Soit A l'événement "la somme des résultats vaut 5". On a $A = \{14, 23, 32, 14\}$ et $\mathbb{P}(A) = \frac{4}{36} = \frac{1}{9}$. Soient X_1 le résultat du dé rouge, X_2 le résultat du dé bleu et S la somme. Ce sont des variables aléatoires et on a $X_1(i_1 i_2) = i_1$, $X_2(i_1 i_2) = i_2$, $S(i_1 i_2) = i_1 + i_2 = X_1(i_1 i_2) + X_2(i_1 i_2)$. Il est immédiat que, pour $k = 1, \ldots, 6$, $\mathbb{P}(X_1 = k) = \mathbb{P}(X_2 = k) = \frac{1}{6}$. La loi de X_1 (et de X_2) est donc la loi uniforme sur $\{1, 2, 3, 4, 5, 6\}$. Soit $(q_k,\ k = 2, 3, \ldots, 12)$ la loi de S. Ci-dessus, on a calculé q_5. De la même façon, on obtient:

$$q_2 = q_{12} = \frac{1}{36},\ q_3 = q_{11} = \frac{2}{36},\ q_4 = q_{10} = \frac{3}{36},\ q_5 = q_9 = \frac{4}{36},\ q_6 = q_8 = \frac{5}{36},\ q_7 = \frac{6}{36}.$$

3. On met au hasard trois boules distinctes a, b, c dans trois urnes. L'ensemble des issues possibles est

$$\Omega = \{(abc| - |-), (-|abc|-), (-| - |abc), (ab|c|-), \ldots\ldots\}.$$

On a $|\Omega| = 3^3 = 27$ et, les issues étant équiprobables, $\mathbb{P}(\{\omega\}) = \frac{1}{27}$. Soit A l'événement "la première urne contient deux boules, la seconde une boule", événement qu'on note $(2|1|0)$. On a $A = \{(ab|c|-), (ac|b|-), (bc|a|-)\}$ d'où $\mathbb{P}(A) = \frac{3}{27} = \frac{1}{9}$. Soit B l'événement "chaque urne contient une boule", événement qu'on note $(1|1|1)$. On a $B = \{(a|b|c), (b|a|c), (a|c|b), (c|a|b), (b|c|a), (c|b|a)\}$ et $\mathbb{P}(B) = \frac{6}{27} = \frac{2}{9}$. Par symétrie, on a

$$\mathbb{P}((3|0|0)) = \mathbb{P}((0|3|0)) = \mathbb{P}((0|0|3)) = \frac{1}{27},$$

$$\mathbb{P}((2|1|0)) = \mathbb{P}((1|2|0)) = \mathbb{P}((2|0|1)) = \mathbb{P}((1|0|2)) = \mathbb{P}((0|2|1)) = \mathbb{P}((0|1|2)) = \frac{1}{9},$$

$$\mathbb{P}((1|1|1)) = \frac{2}{9}.$$

4. On met au hasard trois boules indistinctes dans trois urnes. L'ensemble des issues possibles est

$$\Omega = \{(3|0|0), (0|3|0), (0|0|3), (2|1|0), (1|2|0), (2|0|1), (1|0|2), (0|2|1), (0|1|2), (1|1|1)\}.$$

Mais, vu l'exemple précédent, Ω doit être muni de la probabilité

$$(\frac{1}{27}, \frac{1}{27}, \frac{1}{27}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{27})$$

et non de la probabilité uniforme. Bien sur, Ω muni de la probabilité uniforme est un espace de probabilité mais il ne rend pas compte de l'expérience aléatoire considérée.

1.2. Echantillon. Sous population

Soit $S = \{s_1, s_2, \ldots, s_n\}$ une population de taille n.

1.2.1. Echantillon sans répétition. On tire un par un et sans remise r éléments de S, $r \leq n$. On obtient ce qu'on appelle un échantillon sans répétition de taille r de la population S. C'est une suite $s_{i_1} s_{i_2} \ldots s_{i_r}$ d'éléments de S tous distincts. L'ensemble des issues possibles est donc

$$\Omega = \{s_{i_1} s_{i_2} \ldots s_{i_r},\ s_{i_j} \in S,\ s_{i_j} \neq s_{i_k} \text{ si } j \neq k\}.$$

On a

$$|\Omega| = n(n-1)\ldots(n-r+1) = \frac{n!}{(n-r)!} = A_n^r.$$

$|\Omega|$ est le nombre d'applications injectives de $\{1, 2, \ldots, r\}$ dans $\{1, 2, \ldots, n\}$. Evidemment chaque échantillon a la même probabilité et

$$\mathbb{P}(\{\omega\}) = \frac{1}{|\Omega|} = \frac{(n-r)!}{n!}.$$

Exemple. On suppose $S = \{1, 2, 3, 4\}$ et $r = 2$. Alors $|\Omega| = 12$ et

$$\Omega = \{12, 13, 14, 21, 23, 24, 31, 32, 34, 41, 42, 43\}.$$

1.2.2. Echantillon avec répétitions. On tire un par un et avec remise r éléments de S, r quelconque. On obtient ce qu'on appelle un échantillon avec répétition de taille r de la population S. C'est une suite $s_{i_1} s_{i_2} \ldots s_{i_r}$ d'éléments de S. L'ensemble des issues possibles est donc

$$\Omega = \{s_{i_1} s_{i_2} \ldots s_{i_r},\ s_{i_j} \in S\}.$$

On a

$$|\Omega| = n^r.$$

$|\Omega|$ est le nombre d'applications de $\{1, 2, \ldots, r\}$ dans $\{1, 2, \ldots, n\}$. Evidemment chaque échantillon a la même probabilité et

$$\mathbb{P}(\{\omega\}) = \frac{1}{|\Omega|} = \frac{1}{n^r}.$$

Exemple. On suppose $S = \{1, 2, 3, 4\}$ et $r = 2$. Alors $|\Omega| = 16$ et

$$\Omega = \{11, 12, 13, 14, 21, 22, 23, 24, 31, 32, 33, 34, 41, 42, 43, 44\}.$$

1.2.3. Sous population. On tire en une fois r éléments de S, $r \le n$. On obtient ce qu'on appelle une sous population de taille r de S. C'est un sous ensemble $\{s_{i_1}, s_{i_2}, \ldots, s_{i_r}\}$ de r éléments de S nécessairement distincts (l'ordre n'intervient pas) qu'on écrira simplement $s_{i_1}s_{i_2}\ldots s_{i_r}$. L'ensemble des issues possibles est donc

$$\Omega = \{s_{i_1}s_{i_2}\ldots s_{i_r},\ s_{i_j} \in S,\ i_1 < i_2 < \ldots < i_r\}.$$

On a

$$|\Omega| = C_n^r = \frac{n!}{r!(n-r)!}.$$

$|\Omega|$ est le nombre de sous-ensembles à r éléments d'un ensemble à n éléments. Evidemment chaque sous population a la même probabilité et

$$\mathbb{P}(\{\omega\}) = \frac{1}{|\Omega|} = \frac{r!(n-r)!}{n!}.$$

Exemple. On suppose $S = \{1, 2, 3, 4\}$ et $r = 2$. Alors $|\Omega| = 6$ et

$$\Omega = \{12, 13, 14, 23, 24, 34\}.$$

1.2.4. Loi hypergéométrique. On suppose que $S = S_1 \cup S_2$ avec $S_1 \cap S_2 = \emptyset$, $|S_1| = n_1$, $|S_2| = n_2$, $n = n_1 + n_2$. On appelle éléments de type 1 les éléments de S_1, éléments de type 2 ceux de S_2. On tire sans remise r éléments de S ($r \le n$). Soit X le nombre d'éléments de type 1 obtenus. On se place dans le cadre de 1.2.1 et il s'agit de calculer la loi de la v.a. X. On doit calculer $|A|$ où $A = \{X = k\}$. Evidemment $\mathbb{P}(A) = 0$ si $k > n_1$ ou si $r - k > n_2$. Sinon on construit un élément de A en se donnant un échantillon sans répétition de taille k de S_1 (il y en a $A_{n_1}^k$) puis en se donnant un échantillon sans répétition de taille $r - k$ de S_2 (il y en a $A_{n_2}^{r-k}$) et en faisant un échantillon sans répétition de taille r de S i.e en choisissant la place des éléments de S_1 dans l'échantillon total (il y a donc C_r^k possibilités). Finalement $|A| = A_{n_1}^k A_{n_2}^{r-k} C_r^k$ et

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{n_1!}{(n_1-k)!}\frac{n_2!}{(n_2-(r-k))!}\frac{r!}{k!(r-k)!}\frac{(n-r)!}{n!} = \frac{C_{n_1}^k C_{n_2}^{r-k}}{C_n^r}.$$

En fait il est plus simple de se placer dans le cadre de 1.2.3 et de supposer qu'on tire une sous population de taille r. On a alors $A = \{X = k\} = \{$sous population de taille k de S_1, sous population de taille $r - k$ de $S_2\}$ et $|A| = C_{n_1}^k C_{n_2}^{r-k}$ d'où

$$\mathbb{P}(X = k) = \frac{C_{n_1}^k C_{n_2}^{r-k}}{C_n^r} \text{ convenant que } C_j^i = 0 \text{ si } i > j. \tag{1.6}$$

Cette loi s'appelle la loi hypergéométrique.

1.2.5. Loi binomiale. On suppose encore que $S = S_1 \cup S_2$ avec $S_1 \cup S_2 = \emptyset$, $|S_1| = n_1$, $|S_2| = n_2$, $n = n_1 + n_2$. On tire avec remise r éléments de S, r quelconque, et soit X le nombre d'éléments de type 1 obtenus. On se place dans le cadre de 1.2.2 et il s'agit de calculer la loi de la v.a. X. On doit calculer $|A|$ où $A = \{X = k\}$. Evidemment $\mathbb{P}(A) = 0$ si $k > r$. Sinon on construit un élément de A en se donnant un échantillon avec répétition de taille k de S_1 (il y en a n_1^k) puis en se donnant un échantillon avec répétition de taille $r - k$ de S_2 (il y en a n_2^{r-k}) et en faisant un échantillon avec répétition de taille r de S i.e en choisissant la place des éléments de S_1 dans l'échantillon total (il y a donc C_r^k possibilités). Ceci donne $|A| = n_1^k n_2^{r-k} C_r^k$ et

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = n_1^k n_2^{r-k} C_r^k / n^r.$$

Posant $p = n_1/n$, on obtient

$$\mathbb{P}(X = k) = C_r^k p^k (1-p)^{r-k},\ k = 0, 1, \ldots, r,\ \ \mathbb{P}(X = k) = 0 \text{ si } k > r. \qquad (1.7)$$

Cette loi s'appelle la loi binomiale car $1 = \sum_{k=0}^r \mathbb{P}(X = k)$ n'est rien d'autre que la formule du binôme $\sum_{k=0}^r C_r^k p^k (1-p)^{r-k} = (p + (1-p))^r = 1$.

Evidemment si n_1 et n_2 sont très grands par rapport à r, le fait de tirer sans remise ou avec remise modifie peu le résultat et dans ce cas la loi binomiale est une bonne approximation de la loi hypergéométrique. C'est ce que montre le calcul suivant où k, r sont fixes et où $n_1, n_2 \to +\infty$ avec $n_1/n \to p$. Alors

$$\frac{C_{n_1}^k C_{n_2}^{r-k}}{C_n^r} = \frac{r!\, n_1(n_1-1)\ldots(n_1-k+1) n_2(n_2-1)\ldots(n_2-r+k+1)}{n(n-1)\ldots(n-r+1)\, k!(r-k)!}$$
$$\sim C_r^k \frac{n_1^k n_2^{r-k}}{n^r} = C_r^k (\frac{n_1}{n})^k (1 - \frac{n_1}{n})^{r-k} \to C_r^k p^k (1-p)^{r-k}.$$

1.2.6. Généralisation. On suppose maintenant que $S = S_1 \cup S_2 \cup \ldots \cup S_m$ avec les S_j deux à deux disjoints, $|S_j| = n_j$, $n = n_1 + \ldots + n_m$. On appelle éléments de type j les éléments de S_j, $j = 1, \ldots, m$. On tire sans remise (resp. avec remise) r éléments de S ($r \leq n$ dans le premier cas) et soit X_j le nombre d'éléments de type j obtenus. On veut calculer $\mathbb{P}(X_1 = k_1, \ldots, X_m = k_m)$, $k_1 + \ldots + k_m = r$, on a

a. Tirage sans remise.

$$\mathbb{P}(X_1 = k_1, \ldots, X_m = k_m) = \frac{C_{n_1}^{k_1} \ldots C_{n_m}^{k_m}}{C_n^r},\ \forall j,\ k_j \leq n_j,\ k_1 + \ldots k_m = r\,;\ \ = 0 \text{ sinon.}$$

b. Tirage avec remise. On pose $p_j = \frac{n_j}{n}$. Alors

$$\mathbb{P}(X_1 = k_1, \ldots, X_m = k_m) = \frac{r!}{k_1! \ldots k_m!} p_1^{k_1} \ldots p_m^{k_m},\ k_1 + \ldots k_m = r\,;\ \ = 0 \text{ sinon.}$$

Si $m = 2$, il s'agit des formules précédentes. Dans le cas général, elles se montrent de la même façon.

Exemple. Le bridge se joue avec un jeu de 52 cartes de 4 couleurs. Il oppose deux camps de chacun deux joueurs. On distribue 13 cartes à chaque joueur. On dit qu'une main est 5521 si elle se compose de deux couleurs de 5 cartes, d'une couleur de 2 cartes et d'une couleur de 1 carte. Quelle est la probabilité p qu'une main soit 5521? La probabilité pour qu'une main comprenne 5 piques, 5 cœurs, 2 carreaux, 1 tréfle est (loi hypergéométrique généralisée)

$$\alpha = \frac{C_{13}^5 C_{13}^5 C_{13}^2 C_{13}^1}{C_{52}^{13}} = 0,002645.$$

On obtient la probabilité cherchée en permutant les couleurs. Il y a C_4^2 façons de choisir les deux couleurs de 5 cartes puis deux façons de choisir la couleur de 2 cartes. On a donc $p = 2C_4^2\alpha = 0,03174$.

Vous jouez un contrat avec pique comme atout. Vous avez avec votre partenaire (le mort) 9 piques . Quelles sont les probabilités q_1, q_2, q_3 que, chez vos adversaires, les piques soient partagés $4-0$, $3-1$, $2-2$? La probabilité qu'un de vos adversaires ait 4 (resp. 3, resp. 2) piques est (loi hypergéométrique)

$$\frac{C_4^4 C_{22}^9}{C_{26}^{13}} = 0,0478, \quad \text{resp. } \frac{C_4^3 C_{22}^{10}}{C_{26}^{13}} = 0,2486, \quad \text{resp. } \frac{C_4^2 C_{22}^{11}}{C_{26}^{13}} = 0,40695.$$

On a donc $q_1 = 0,09565$, $q_2 = 0,4974$, $q_3 = 0,40695$.

1.3. Probabilité conditionnelle

On considère un espace de probabilité fini $(\Omega, \mathbb{P})$. On écrit indifféremment $A \cap B$ ou AB.

1.3.1. Probabilité conditionnelle.

Soient Ω une population, A la sous population des hommes, A^c celle des femmes et B celle des fumeurs. Si on tire au hasard un élément de Ω, la probabilité d'obtenir un fumeur est $\frac{|B|}{|\Omega|}$. Si on observe que l'élément tiré est un homme, la probabilité que ce soit un fumeur est $\frac{|AB|}{|A|}$, c'est ce qu'on appellera la probabilité conditionnelle de B sachant A. Ceci conduit à:

Définition 1.3.1. *Soit $A \subset \Omega$ tel que $\mathbb{P}(A) > 0$. On appelle probabilité conditionnelle de B sachant A et on note $\mathbb{P}(B|A)$ la quantité $\mathbb{P}(AB)/\mathbb{P}(A)$.*

On a donc

$$\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B|A).$$

Noter que $B \mapsto \mathbb{P}(B|A)$ est une probabilité sur Ω.

Proposition 1.3.2. *(Formule de Bayes) Soient* A, B *des événements tels que* $\mathbb{P}(A) > 0, \mathbb{P}(A^c) > 0, \mathbb{P}(B) > 0$. *On a*

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A)\mathbb{P}(B|A)}{\mathbb{P}(A)\mathbb{P}(B|A) + \mathbb{P}(A^c)\mathbb{P}(B|A^c)}\,.$$

Preuve: Par définition $\mathbb{P}(A|B) = \mathbb{P}(AB)/\mathbb{P}(B)$. D'une part $\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B|A)$. D'autre part $\mathbb{P}(B) = \mathbb{P}(BA) + \mathbb{P}(BA^c) = \mathbb{P}(A)\mathbb{P}(B|A) + \mathbb{P}(A^c)\mathbb{P}(B|A^c)$. D'où le résultat. $\diamond$

Proposition 1.3.3. *Soient* $A_1, A_2, \ldots, A_n$ *des événements tels que* $\mathbb{P}(A_1A_2\ldots A_n) > 0$. *On a*

$$\mathbb{P}(A_1A_2\ldots A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1A_2)\ldots\mathbb{P}(A_n|A_1A_2\ldots A_{n-1})\,.$$

Preuve: Par définition $\mathbb{P}(A_1A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)$. Supposons la formule vraie au rang n. Alors $\mathbb{P}(A_1A_2\ldots A_nA_{n+1}) = \mathbb{P}(A_1A_2\ldots A_n)\mathbb{P}(A_{n+1}|A_1A_2\ldots A_n)$ et il suffit d'appliquer la formule au rang n pour conclure. $\diamond$

1.3.2. Evénements indépendants. Si $\mathbb{P}(B|A) = \mathbb{P}(B)$ i.e. $\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B)$, savoir si A a eu lieu ou non ne modifie pas la probabilité de B. Il est alors naturel de dire que les événements A et B sont indépendants d'où

Définition 1.3.4. *Les événements* A *et* B *sont indépendants si* $\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B)$.

Supposons A et B indépendants, on a

$$\mathbb{P}(AB^c) = \mathbb{P}(A) - \mathbb{P}(AB) = \mathbb{P}(A) - \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A)(1 - \mathbb{P}(B)) = \mathbb{P}(A)\mathbb{P}(B^c).$$

Donc A et B^c sont indépendants. On voit facilement qu'il en est de même de A^c et B et de A^c et B^c. Donc posant, pour $F \subset \Omega$;

$$\sigma(F) = \{\Omega, F, F^c, \emptyset\}, \tag{1.8}$$

on a que A et B sont indépendants ssi $\mathbb{P}(CD) = \mathbb{P}(C)\mathbb{P}(D)$ pour tout $C \in \sigma(A)$ et tout $D \in \sigma(B)$. Ceci conduit à:

Définition 1.3.5. *Les événements* $A_1, A_2, \ldots, A_n$ *sont indépendants si, pour tout* $C_1 \in \sigma(A_1)$, *tout* $C_2 \in \sigma(A_2)$,..., *tout* $C_n \in \sigma(A_n)$,

$$\mathbb{P}(C_1C_2\ldots C_n) = \mathbb{P}(C_1)\mathbb{P}(C_2)\ldots\mathbb{P}(C_n).$$

On montre alors facilement:

Proposition 1.3.6. *Les événements* $A_1, A_2, \ldots, A_n$ *sont indépendants ssi, pour tout* $\{i_1, \ldots, i_k\} \subset \{1, \ldots, n\}$,

$$\mathbb{P}(A_{i_1}\ldots A_{i_k}) = \mathbb{P}(A_{i_1})\ldots\mathbb{P}(A_{i_k}).$$

Chapitre 2

Espace de probabilité discret

Dans ce chapitre, on introduit les espaces de probabilité dénombrables. Pour cela, on a besoin de la notion de famille sommable.

2.1. Famille sommable

Dans toute cette section, I désigne un ensemble dénombrable.

2.1.1. Notations. Soient E un ensemble, $A_n \subset E$ et $f_n : E \to \mathbb{R}$. On écrit $A_n \uparrow A$ si $A_n \subset A_{n+1}$ et $A = \cup A_n$, $A_n \downarrow A$ si $A_n \supset A_{n+1}$ et $A = \cap A_n$, $f_n \uparrow f$ si $f_n \leq f_{n+1}$ et $f = \sup f_n$ (alors $f = \lim \uparrow f_n$), $f_n \downarrow f$ si $f_n \geq f_{n+1}$ et $f = \inf f_n$ (alors $f = \lim \downarrow f_n$).

2.1.2. Enumération. On appelle énumération de I toute bijection ϕ de $\mathbb{N}$ sur I. Soient $(a_i,\, i \in I)$ une famille de nombres réels ou complexes et ϕ une énumération de I. On pose

$$S_n^\phi = a_{\phi(0)} + a_{\phi(1)} + \ldots + a_{\phi(n)}. \tag{2.1}$$

2.1.3. Famille sommable positive. On suppose que, pour tout $i \in I$, $a_i \geq 0$. Alors la suite S_n^ϕ est croissante. Soit $S^\phi = \lim \uparrow S_n^\phi \in \overline{\mathbb{R}}^+$. Si ψ est une autre énumération de I, on a, pour n fixé et m assez grand,

$$\{a_{\phi(0)}, a_{\phi(1)}, \ldots, a_{\phi(n)}\} \subset \{a_{\psi(0)}, a_{\psi(1)}, \ldots, a_{\psi(m)}\}$$

et donc $S_n^\phi \leq S_m^\psi \leq S^\psi$ d'où $S^\phi \leq S^\psi$. Changeant le rôle de ϕ et ψ, on a $S^\psi \leq S^\phi$ et finalement $S^\phi = S^\psi$. On peut énoncer:

Théorème 2.1.1. *Soit $(a_i,\, i \in I)$ une famille de nombres positifs. Alors, pour toute énumération ϕ de I, la suite S_n^ϕ, définie par (2.1), converge en croissant vers un nombre $S \in \overline{\mathbb{R}}^+$ indépendant de ϕ. On note $S = \sum_{i\in I} a_i$. Si $S < +\infty$, la famille est dite sommable.*

Quelques conséquences immédiates:

(i) Si $I_n \uparrow I$, I_n fini, $\sum_{i\in I_n} a_i \uparrow \sum_{i\in I} a_i$.

(ii) Pour tout $A < \sum_{i\in I} a_i$, il existe $J \subset I$, J fini, tel que $\sum_{i\in J} a_i > A$.

(iii) Si $0 \le a_i \le b_i$, $\sum_{i\in I} a_i \le \sum_{i\in I} b_i$.

(iv) Pour $\alpha \ge 0$, $\beta \ge 0$, $a_i \ge 0$, $b_i \ge 0$, on a

$$\sum_{i\in I}(\alpha a_i + \beta b_i) = \alpha \sum_{i\in I} a_i + \beta \sum_{i\in I} b_i.$$

Remarque. En fait $\sum_{i\in I} a_i$ est défini pour $a_i \in \overline{\mathbb{R}}^+$ et vaut $+\infty$ si $a_i = +\infty$ pour un i au moins.

2.1.4. Passage à la limite croissante.

Proposition 2.1.2. *Soit, pour tout $n \in \mathbb{N}$, $(a_i(n),\, i \in I)$ une famille de réels positifs. On suppose que, pour tout $i \in I$, $a_i(n) \uparrow a_i$ lorsque $n \to +\infty$. Alors*

$$\sum_{i\in I} a_i(n) \uparrow \sum_{i\in I} a_i \ \textit{lorsque}\ n \to +\infty.$$

Preuve: Soient $S(n) = \sum_{i\in I} a_i(n)$, $S^* = \lim \uparrow_n S(n)$, $S = \sum_{i\in I} a_i$. Evidemment $S^* \le S$. Soit $A < S$. Il existe J fini, $J \subset I$, tel que $\sum_{i\in J} a_i > A$. Donc, pour n assez grand, $\sum_{i\in J} a_i(n) > A$ et $S^* \ge A$ d'où $S^* \ge S$ et $S^* = S$. $\diamond$

2.1.5. Sommation par paquets. On dit que $(I_j,\, j \in J)$ est une partition de I si les I_j sont deux à deux disjoints et si $I = \cup_{j\in J} I_j$.

Proposition 2.1.3. *Soient $(a_i,\, i \in I)$ une famille de réels positifs et $(I_j,\, j \in J)$ une partition de I. On a*

$$\sum_{i\in I} a_i = \sum_{j\in J}\sum_{i\in I_j} a_i.$$

Preuve: Soient $K_n \uparrow I$, K_n fini, et $J_n = \{j \in J,\, K_n \cap I_j \ne \emptyset\}$. K_n et J_n étant finis,

$$\sum_{i\in K_n} a_i = \sum_{j\in J_n}\sum_{i\in I_j\cap K_n} a_i = \sum_{j\in J} b_j(n)$$

où $b_j(n) = 0$ si $j \notin J_n$, $b_j(n) = \sum_{i\in I_j\cap K_n} a_i$ si $j \in J_n$. D'une part $\sum_{i\in K_n} a_i \uparrow_n \sum_{i\in I} a_i$ et d'autre part, pour chaque j, $b_j(n) \uparrow_n \sum_{i\in I_j} a_i$ d'où (prop. 2.1.2) $\sum_{j\in J} b_j(n) \uparrow_n \sum_{j\in J}\sum_{i\in I_j} a_i$. $\diamond$

2.1.6. Le cas général. On considère maintenant une famille $(a_i,\, i \in I)$ de nombres réels ou complexes.

Définition 2.1.4. *Une famille $(a_i,\, i \in I)$ de nombres réels ou complexes est dite sommable si $\sum_{i\in I} |a_i| < +\infty$.*

Théorème 2.1.5. *Soit*$(a_i,\, i \in I)$ *une famille sommable de nombres complexes.*

(i) Pour toute énumération ϕ *de* I, S_n^ϕ *définie par (2.1) converge vers* $S \in \mathbb{C}$ *indépendant de* ϕ. *On note* $S = \sum_{i\in I} a_i$. *On a* $|\sum_{i\in I} a_i| \leq \sum_{i\in I} |a_i|$.

(ii) Soit $(I_j,\, j \in J)$ *une partition de* I, *on a* $\sum_{i\in I} a_i = \sum_{j\in J}\sum_{i\in I_j} a_i$.

(iii) Si $(b_i,\, i \in I)$ *est une autre famille sommable de nombres complexes et si* $\alpha, \beta \in \mathbb{C}$, *la famille* $(\alpha a_i + \beta b_i,\, i \in I)$ *est sommable et*

$$\sum_{i\in I}(\alpha a_i + \beta b_i) = \alpha \sum_{i\in I} a_i + \beta \sum_{i\in I} b_i.$$

Preuve: On pose, pour $a \in \mathbb{R}$, $a^+ = \max(a, 0)$, $a^- = \max(-a, 0)$. On a $a = a^+ - a^-$ et $|a| = a^+ + a^-$. Pour $a \in \mathbb{C}$, on a $a = \Re(a) + \mathtt{i}\Im(a)$. Alors, pour tout $i \in I$,

$$[\Re(a_i)]^+ \leq |a_i|,\ \ [\Re(a_i)]^- \leq |a_i|,\ \ [\Im(a_i)]^+ \leq |a_i|,\ \ [\Im(a_i)]^- \leq |a_i|.$$

Ecrivant

$$S_n^\phi = \sum_{k=0}^n [\Re(a_{\phi(k)})]^+ - \sum_{k=0}^n [\Re(a_{\phi(k)})]^- + \mathtt{i}\sum_{k=0}^n [\Im(a_{\phi(k)})]^+ - \mathtt{i}\sum_{k=0}^n [\Im(a_{\phi(k)})]^-,$$

on est ramené au cas positif. $\diamond$

2.2. Espace de probabilité discret

2.2.1. Probabilité sur E dénombrable. Soit E un ensemble dénombrable. Une probabilité sur E est une famille $(p(a),\, a \in E)$ de réels vérifiant

$$0 \leq p(a) \leq 1, \quad \sum_{a\in E} p(a) = 1.$$

On pose alors, pour $A \subset E$, $\mathbb{P}(A) = \sum_{a\in A} p(a)$. $\mathbb{P}$ est une application de $\mathcal{P}(E)$ dans $[0, 1]$ vérifiant $\mathbb{P}(E) = 1$, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ si $A \cap B = \emptyset$ (prop. 2.1.3) et $\mathbb{P}(A_n) \uparrow \mathbb{P}(A)$ si $A_n \uparrow A$ (prop. 2.1.2). Ceci implique que $A \mapsto \mathbb{P}(A)$ est σ-additive i.e. que, pour toute famille $(A_n,\, n \in \mathbb{N})$ de sous-ensembles de Ω deux à deux disjoints, on a $\mathbb{P}(\cup A_n) = \sum \mathbb{P}(A_n)$. En effet:

$$\mathbb{P}(\cup A_n) = \lim \uparrow_N \mathbb{P}(\cup_0^N A_n) = \lim \uparrow_N \sum_0^N \mathbb{P}(A_n) = \sum \mathbb{P}(A_n).$$

Réciproquement si une application de $\mathcal{P}(E)$ dans $[0, 1]$, $A \mapsto \mathbb{P}(A)$, vérifie $\mathbb{P}(E) = 1$ et est σ-additive, on a, posant $p(a) = \mathbb{P}(\{a\})$, $0 \leq p(a) \leq 1$ et $\sum_{a\in E} p(a) = 1$. Ici encore, on appellera probabilité sur E aussi bien la famille $(p(a),\, a \in E)$ que la fonction d'ensembles $A \mapsto \mathbb{P}(A)$.

2.2.2. Un couple $(\Omega, \mathbb{P})$ où Ω est un ensemble fini ou dénombrable et $\mathbb{P}$ une probabilité sur Ω s'appelle un espace de probabilité discret. Toute application X de Ω dans E s'appelle une variable aléatoire à valeurs E. On peut supposer E dénombrable puisque $X(\Omega)$ est dénombrable. Alors, vu la prop. 2.1.3, la famille $(q(a),\ a \in E)$ où $q(a) = \mathbb{P}(X = a)$ est une probabilité sur E appelée loi de X.

2.2.3. Espérance. Soient $(\Omega, \mathbb{P})$ un espace de probabilité discret et X une variable aléatoire à valeurs E discret (i.e. fini ou dénombrable). On pose $p(\omega) = \mathbb{P}(\{\omega\})$.

a. On suppose $E \subset \mathbb{R}^+$. On pose $\mathbb{E}(X) = \sum_{\omega\in\Omega} X(\omega)p(\omega)$. $\mathbb{E}(X)$, qui est un élément de $[0, +\infty]$, s'appelle l'espérance de X.

b. On suppose $E \subset \mathbb{R}$. Alors, si $\mathbb{E}(|X|) = \sum_\omega |X(\omega)|p(\omega) < +\infty$, on appelle espérance de X la quantité $\mathbb{E}(X) = \sum_{\omega\in\Omega} X(\omega)p(\omega)$.

c. On suppose E quelconque et soit $f : E \to \mathbb{R}$. Si $f \geq 0$ ou si $\mathbb{E}(|f(X)|) = \sum_{\omega\in\Omega} |f(X(\omega))|p(\omega) < +\infty$, on a

$$\mathbb{E}(f(X)) = \sum_{\omega\in\Omega} f(X(\omega))p(\omega). \tag{2.2}$$

Théorème 2.2.1. *Soient X une variable aléatoire à valeurs E discret et $f : E \to \mathbb{R}$. Si $f \geq 0$, on a*

$$\mathbb{E}(f(X)) = \sum_{a\in E} f(a)\mathbb{P}(X = a). \tag{2.3}$$

De plus, $\mathbb{E}(|f(X)|) < +\infty$ ssi $\sum_a |f(a)|\mathbb{P}(X = a) < +\infty$ et, dans ce cas, on a (2.3).

Preuve: Supposons d'abord $f \geq 0$. Alors, vu la prop. 2.1.3,

$$\mathbb{E}(f(X)) = \sum_{\omega\in\Omega} f(X(\omega))p(\omega) = \sum_{a\in E} \sum_{\omega \,/\, X(\omega)=a} f(X(\omega))p(\omega)$$
$$= \sum_{a\in E} \sum_{\omega \,/\, X(\omega)=a} f(a)p(\omega) = \sum_{a\in E} f(a) \sum_{\omega \,/\, X(\omega)=a} p(\omega) = \sum_{a\in E} f(a)\mathbb{P}(X = a).$$

On a donc, pour f réelle, $\mathbb{E}(|f(X)|) = \sum_a |f(a)|\mathbb{P}(X = a)$ et, si cette quantité est finie, le calcul ci dessus est encore valable (th. 2.1.5). $\diamond$

Soient X_1, X_2 des v.a. à valeurs E_1 et E_2 discrets. Alors (X_1, X_2) est une v.a. à valeurs $E_1 \times E_2$ et on a, pour toute $f : E_1 \times E_2 \to \mathbb{R}$ positive ou telle que $\mathbb{E}(|f(X_1, X_2)|) < +\infty$,

$$\mathbb{E}(f(X_1, X_2)) = \sum_{(a_1,a_2)\in E_1\times E_2} f(a_1, a_2)\, \mathbb{P}(X_1 = a_1, X_2 = a_2). \tag{2.4}$$

Si $A \subset \Omega$, on appelle fonction indicatrice de A et on note 1_A la fonction définie par $1_A(\omega) = 1$ si $\omega \in A$, $1_A(\omega) = 0$ si $\omega \notin A$. Alors, notant $p(\omega) = \mathbb{P}(\{\omega\})$,

$$\mathbb{E}(1_A) = \sum_{\omega\in\Omega} 1_A(\omega)p(\omega) = \sum_{\omega\in A} p(\omega) = \mathbb{P}(A). \tag{2.5}$$

2.2.4. Moments. Dans cette sous section, X désigne une v.a. à valeurs $E \subset \mathbb{R}$, E discret. Soit $p \in \mathbb{N}^*$. Si $\mathbb{E}(|X|^p) < +\infty$, $\mathbb{E}(|X|^p)$ s'appelle le moment absolu d'ordre p de X et $\mathbb{E}(X^p)$ s'appelle le moment d'ordre p de X. D'après le th. 2.2.1,

$$\mathbb{E}(|X|^p) = \sum_{a \in E} |a|^p \, \mathbb{P}(X = a).$$

Noter que, pour $1 \leq q \leq p$, $\mathbb{E}(|X|^p) < +\infty$ implique $\mathbb{E}(|X|^q) < +\infty$ puisque $|X|^q \leq 1 + |X|^p$.

Supposons $\mathbb{E}(X^2) < +\infty$, alors $m = \mathbb{E}(X)$, qu'on appelle aussi moyenne de X, existe et on définit la variance de X par

$$\mathrm{Var}(X) = \mathbb{E}[(X - m)^2] = \mathbb{E}(X^2) - m^2. \tag{2.6}$$

La variance donne une idée de l'écart de X par rapport à sa moyenne m comme le montre:

Proposition 2.2.2. *(Inégalité de Bienaymé-Tchebychev) On suppose que* $\mathbb{E}(X^2) < +\infty$ *et soit* $m = \mathbb{E}(X)$*. Alors, pour tout* $\lambda > 0$*,*

$$\mathbb{P}(|X - m| \geq \lambda) \leq \frac{1}{\lambda^2} \, Var(X).$$

Preuve: On a

$$\mathrm{Var}(X) = \mathbb{E}[(X-m)^2] = \sum_{\omega \in \Omega} (X(\omega) - m)^2 p(\omega) \geq \sum_{\omega \in \{|X-m| \geq \lambda\}} (X(\omega) - m)^2 p(\omega)$$
$$\geq \lambda^2 \sum_{\omega \in \{|X-m| \geq \lambda\}} p(\omega) = \lambda^2 \, \mathbb{P}(|X - m| \geq \lambda). \diamond$$

2.2.5. Lois usuelles.

Loi binomiale. On l'a déjà rencontré en (1.7). Soit $n \in \mathbb{N}^*$. C'est la loi d'une v.a. à valeurs $\{0, 1, \ldots, n\}$ telle que

$$\mathbb{P}(X = k) = C_n^k p^k (1-p)^{n-k}, \; k = 0, 1, \ldots, n; \;\; 0 < p < 1. \tag{2.7}$$

Elle est appelée loi binomiale de paramètre n, p et notée $B(n, p)$. On écrit $X \sim B(n, p)$. En particulier si $X \sim B(1, p)$, on dit que X est une v.a. de Bernouilli. Calculons la moyenne et la variance de $X \sim B(n, p)$. D'une part

$$\mathbb{E}(X) = \sum_{k \geq 0} k \, \mathbb{P}(X = k) = \sum_{k=1}^{n} k C_n^k p^k (1-p)^{n-k} = np \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!} \, p^{k-1} (1-p)^{n-k}$$
$$= np \sum_{i=0}^{n-1} C_{n-1}^i p^i (1-p)^{n-1-i} = np(p + (1-p))^{n-1} = np.$$

D'autre part

$$
\begin{aligned}
\mathbb{E}(X^2) &= \sum_{k\geq 0} k^2\,\mathbb{P}(X=k) = \sum_{k=2}^{n} k(k-1)C_n^k p^k (1-p)^{n-k} + \sum_{k=1}^{n} k\,\mathbb{P}(X=k)\\
&= n(n-1)p^2 \sum_{k=2}^{n} \frac{(n-2)!}{(k-2)!(n-k)!}\, p^{k-2}(1-p)^{n-k} + pn\\
&= n(n-1)p^2 \sum_{i=0}^{n-2} C_{n-2}^i p^i (1-p)^{n-2-i} + pn = n(n-1)p^2 + pn.
\end{aligned}
$$

On a alors $\mathrm{Var}(X) = n(n-1)p^2 + pn - (np)^2 = np(1-p)$.

Supposons que k soit fixe et que $n \to +\infty$ avec $p = p(n)$ tel que $np(n) \to \lambda$. Alors vu que $\log\{(1-p(n))^n\} = n\log(1-p(n)) \sim -np(n) \to -\lambda$, on a

$$
\begin{aligned}
\mathbb{P}(X=k) &= \frac{n(n-1)\dots(n-k+1)}{k!} p^k(n)(1-p(n))^{n-k}\\
&= \frac{1}{k!}\frac{n(n-1)\dots(n-k+1)}{n^k}(np(n))^k(1-p(n))^{-k}(1-p(n))^n \to \frac{1}{k!}\lambda^k \mathrm{e}^{-\lambda}.
\end{aligned}
$$

Noter que $(\frac{1}{k!}\lambda^k \mathrm{e}^{-\lambda},\ k\in\mathbb{N})$ est une probabilité sur $\mathbb{N}$.

Loi de Poisson. C'est la loi d'une v.a. à valeurs $\mathbb{N}$ telle que

$$
\mathbb{P}(X=k) = \mathrm{e}^{-\lambda}\frac{\lambda^k}{k!},\quad k\in\mathbb{N};\quad \lambda>0. \tag{2.8}
$$

Cette loi est appelée loi de Poisson de paramètre λ et se note $\mathcal{P}(\lambda)$. Calculons sa moyenne et sa variance. D'une part

$$
\mathbb{E}(X) = \sum_{k\geq 0} k\,\mathbb{P}(X=k) = \sum_{k=0}^{\infty} k\mathrm{e}^{-\lambda}\frac{\lambda^k}{k!} = \lambda\mathrm{e}^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!} = \lambda.
$$

D'autre part, comme ci-dessus

$$
\begin{aligned}
\mathbb{E}(X^2) &= \sum_{k\geq 0} k^2\,\mathbb{P}(X=k) = \sum_{k\geq 0} k(k-1)\mathrm{e}^{-\lambda}\frac{\lambda^k}{k!} + \sum_{k\geq 0} k\mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}\\
&= \lambda^2\mathrm{e}^{-\lambda}\sum_{k=2}^{\infty}\frac{\lambda^{k-2}}{(k-2)!} + \lambda = \lambda^2+\lambda.
\end{aligned}
$$

On a alors $\mathrm{Var}(X) = \lambda^2 + \lambda - \lambda^2 = \lambda$.

On a vu qu'on peut approximer la loi $B(n,p)$ par la loi de Poisson $\mathcal{P}(np)$ si n est très grand et p très petit.

Loi géométrique. C'est la loi d'une v.a. à valeurs $\mathbb{N}$ telle que

$$
\mathbb{P}(X=k) = (1-a)a^k,\quad k\in\mathbb{N};\quad 0<a<1. \tag{2.9}
$$

Cette loi est appelée loi géométrique sur $\mathbb{N}$ de paramètre a et se note $\mathcal{G}(a)$. On calculera sa moyenne et sa variance en 2.3. On rencontrera aussi la loi géométrique sur $\mathbb{N}^*$ de paramètre a, notée $\mathcal{G}^*(a)$ définie par

$$\mathbb{P}(X = k) = (1-a)a^{k-1}, \quad k \in \mathbb{N}^*, \quad 0 < a < 1. \tag{2.10}$$

2.2.6. Variables aléatoires indépendantes. Il est naturel de dire que deux v.a. discrètes X et Y sont indépendantes si, pour tous $a \in X(\Omega), b \in Y(\Omega)$, les événements $\{X = a\}$ et $\{Y = b\}$ sont indépendants (voir 1.3.2) i.e. si pour tous $a \in X(\Omega), b \in Y(\Omega)$, $\mathbb{P}(X = a, Y = b) = \mathbb{P}(X = a)\mathbb{P}(Y = b)$. Plus généralement,

Définition 2.2.3. *Les v.a. $X_1, X_2, \ldots, X_n$ à valeurs $E_1, E_2, \ldots, E_n$ discrets sont indépendantes si, pour tous $a_1 \in E_1, a_2 \in E_2, \ldots, a_n \in E_n$,*

$$\mathbb{P}(X_1 = a_1, X_2 = a_2, \ldots, X_n = a_n) = \mathbb{P}(X_1 = a_1)\,\mathbb{P}(X_2 = a_2)\ldots\mathbb{P}(X_n = a_n).$$

Théorème 2.2.4. *Les v.a. $X_1, X_2, \ldots, X_n$ à valeurs $E_1, E_2, \ldots, E_n$ discrets sont indépendantes ssi, pour tous $f_i : \mathbb{E}_i \to \mathbb{R}^+$,*

$$\mathbb{E}(f_1(X_1)\ldots f_n(X_n)) = \mathbb{E}(f_1(X_1))\ldots\mathbb{E}(f_n(X_n)) \tag{2.11}$$

Dans ce cas, si $f_i : \mathbb{E}_i \to \mathbb{R}$ vérifie $\mathbb{E}(|f_i(X_i)|) < +\infty$, $i = 1, 2, \ldots, n$, on a que $\mathbb{E}(|f_1(X_1)\ldots f_n(X_n)|) < +\infty$ et (2.11) est satisfaite.

Preuve: On se limite à $n = 2$. Si (2.11) est satisfaite, on a l'indépendance de X_1 et X_2 en choisissant $f_1 = 1_{\{a_1\}}, f_2 = 1_{\{a_2\}}$ et en utilisant (2.5). Réciproquement, si X_1 et X_2 sont indépendantes et $f_1 \geq 0, f_2 \geq 0$, vu la prop. 2.1.3 et (2.4),

$$\begin{aligned}\mathbb{E}(f_1(X_1)f_2(X_2)) &= \sum_{a_1,a_2} f_1(a_1)f_2(a_2)\mathbb{P}(X_1 = a_1, X_2 = a_2)\\ &= \sum_{a_1,a_2} f_1(a_1)f_2(a_2)\mathbb{P}(X_1 = a_1)\mathbb{P}(X_2 = a_2)\\ &= \sum_{a_1} f_1(a_1)\mathbb{P}(X_1 = a_1)\sum_{a_2} f_2(a_2)\mathbb{P}(X_2 = a_2) = \mathbb{E}(f_1(X_1))\mathbb{E}(f_2(X_2)).\end{aligned}$$

Dans le cas réel, on a, vu la première partie, $\mathbb{E}(|f_1(X_1)f_2(X_2)|) = \mathbb{E}(|f_1(X_1)|)\mathbb{E}(|f_2(X_2)|) < +\infty$ et la calcul ci-dessus reste valable. $\diamond$

Prenant $f_i = 1_{\Gamma_i}$, on a, utilisant (2.5), que si $X_1, X_2, \ldots, X_n$ sont indépendantes, pour tous $\Gamma_i \subset E_i$,

$$\mathbb{P}(X_1 \in \Gamma_1, \ldots X_n \in \Gamma_n) = \mathbb{P}(X_1 \in \Gamma_1)\ldots\mathbb{P}(X_n \in \Gamma_n) \tag{2.12}$$

Enfin il résulte du th. 2.2.4 que, si $X_1, X_2, \ldots, X_n$ sont indépendantes,

(i) il en est de même $Y_1 = g_1(X_1), \ldots, Y_n = g_n(X_n)$ où $g_i : E_i \to F_i$.

(ii) il en est de même de $X_{r(1)}, \dots, X_{r(n)}$ pour toute permutation $\{r(1), \dots, r(n)\}$ de $(1, \dots, n)$,

(iii) il en est de même,pour tous $1 < m_1 < \dots < m_p = n$, de $Y_1, \dots, Y_p$ où

$$Y_1 = (X_1, \dots, X_{m_1}), Y_2 = (X_{m_1+1}, \dots, X_{m_2}), \dots, Y_p = (X_{m_{p-1}+1}, \dots, X_n).$$

Par exemple, si X_1, X_2, X_3, X_4 sont des variables aléatoires réelles indépendantes, il en est de même de X_1, X_3, X_2, X_4, de $Y_1 = (X_1, X_3)$ et $Y_2 = (X_2, X_4)$ et de $U_1 = \cos(X_1^2 + X_3^2)$ et $U_2 = \mathrm{e}^{X_2 X_4}$.

Exemple. Soient X et Y deux v.a. indépendantes à valeurs $\mathbb{N}$, de lois $\mathcal{P}(\lambda)$ et $\mathcal{P}(\mu)$. Cherchons la loi de $S = X + Y$. On a

$$\mathbb{P}(S = k) = \mathbb{P}(X + Y = k) = \sum_{j=0}^{k} \mathbb{P}(X = j, Y = k - j) = \sum_{j=0}^{k} \mathbb{P}(X = j)\mathbb{P}(Y = k - j)$$

$$= \sum_{j=0}^{k} \mathrm{e}^{-\lambda} \frac{\lambda^j}{j!} \mathrm{e}^{-\mu} \frac{\mu^{k-j}}{(k-j)!} = \mathrm{e}^{-(\lambda+\mu)} \frac{1}{k!} \sum_{j=0}^{k} C_k^j \lambda^j \mu^{k-j} = \mathrm{e}^{-(\lambda+\mu)} \frac{(\lambda+\mu)^k}{k!}.$$

Donc $S \sim \mathcal{P}(\lambda + \mu)$.

2.3. Fonctions génératrices

Dans cette section, on ne considère que des v.a. à valeurs $\mathbb{N}$.

2.3.1. Définition. Soit X une telle v.a. Notons d'abord que, vu le th. 2.2.1, on a, pour tout $s \geq 0$, $\sum_{n=0}^{\infty} \mathbb{P}(X = n) s^n = \mathbb{E}(s^X)$ avec la convention $s^0 = 1$ si $s = 0$.

Définition 2.3.1. *On appelle fonction génératrice de X, la fonction*

$$g(s) = g_X(s) = \sum_{n=0}^{\infty} \mathbb{P}(X = n) s^n = \mathbb{E}(s^X), \quad 0 \leq s \leq 1.$$

On pose $q_n = \mathbb{P}(X = n)$. On a $g_X(0) = q_0$, $g_X(1) = 1$ et, vu la prop. 2.1.2, $g_X(s) \uparrow g_X(1) = 1$ lorsque $s \uparrow 1$. Sur $[0, 1]$, la fonction $g_X(s)$ est convexe et strictement convexe si $q_0 + q_1 < 1$. De plus, la série entière $\sum q_n s^n$ a un rayon de convergence $R \geq 1$. Donc $g_X(s)$ est indéfiniment dérivable sur $[0, 1[$ et $g'_X(s) = \sum_{n \geq 1} n q_n s^{n-1}$, $g''_X(s) = \sum_{n \geq 2} n(n-1) q_n s^{n-2}$,.... Enfin $n! q_n = g_X^{(n)}(0)$ d'où:

Proposition 2.3.2. *La fonction génératrice g_X détermine la loi de X. En fait:*

$$\mathbb{P}(X = n) = \frac{1}{n!} g_X^{(n)}(0).$$

Exemples.

a. Loi binomiale $B(n,p)$. On a

$$g(s) = \sum_k \mathbb{P}(X=k)s^k = \sum_{k=0}^{n} C_n^k p^k s^k (1-p)^{n-k} = (ps+(1-p))^n.$$

b. Loi de Poisson $\mathcal{P}(\lambda)$. On a

$$g(s) = \sum_k \mathbb{P}(X=k)s^k = \mathrm{e}^{-\lambda} \sum_{k\geq 0} \frac{\lambda^k s^k}{k!} = \mathrm{e}^{\lambda(s-1)}.$$

c. Loi géométrique $\mathcal{G}(a)$. On a

$$g(s) = \sum_k \mathbb{P}(X=)s^k = \sum_{k\geq 0}(1-a)a^k s^k = \frac{1-a}{1-as}.$$

2.3.2. Calcul des moments. Rappelons (2.2.4) que $\mathbb{E}(X^p) < +\infty$ implique $\mathbb{E}(X^q) < +\infty$ pour tout $q \leq p$.

Proposition 2.3.3. *(i)* $\mathbb{E}(X) < +\infty$ *ssi* g_X *est dérivable à gauche en* 1 *et, dans ce cas, on a* $\mathbb{E}(X) = g'_X(1)$.

(ii) $\mathbb{E}(X^2) < +\infty$ *ssi* g_X *est deux fois dérivable à gauche en* 1 *et, dans ce cas, on a* $\mathbb{E}(X(X-1)) = g''_X(1)$.

Preuve: (i) On a, utilisant la prop. 2.1.2, lorsque $s \uparrow 1$,

$$\frac{g(s)-g(1)}{s-1} = \sum_{n\geq 0} q_n \frac{s^n-1}{s-1} = \sum_{n\geq 0} q_n(1+\ldots+s^{n-1}) \uparrow \sum_{n\geq 0} nq_n = \mathbb{E}(X)$$

et le résultat cherché.

(ii) On remarque d'abord que, si $\mathbb{E}(X^2) < +\infty$, $\mathbb{E}(X) < +\infty$ et $g'(1) < +\infty$. Alors, lorsque $s \uparrow 1$,

$$\frac{g'(s)-g'(1)}{s-1} = \sum_{n\geq 0} nq_n \frac{s^{n-1}-1}{s-1} = \sum_{n\geq 0} nq_n(1+\ldots+s^{n-2}) \uparrow \sum_{n\geq 0} n(n-1)q_n = \mathbb{E}(X(X-1)).$$

On conclut facilement. $\diamond$

On peut continuer et, si $\mathbb{E}(X^p) < +\infty$, $p \in \mathbb{N}$,

$$g_X^{(p)}(1) = \mathbb{E}(X(X-1)\ldots(X-p+1)).$$

Supposons $\mathbb{E}(X^2) < +\infty$. Alors

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = \mathbb{E}(X(X-1)) + \mathbb{E}(X) - [\mathbb{E}(X)]^2 = g''_X(1) + g'_X(1) - [g'_X(1)]^2.$$

Le lecteur est invité à calculer l'espérance et la variance des lois binomiale et de Poisson par cette méthode. Considérons la loi géométrique $\mathcal{G}(a)$ (2.3.1). On a

$$g(s) = \frac{1-a}{1-as},\ g'(1) = \frac{a}{1-a} = \mathbb{E}(X),\ g''(1) = \frac{2a^2}{(1-a)^2},\ \mathrm{Var}(X) = \frac{a}{(1-a)^2}.$$

2.3.3. Somme de v.a. indépendantes.

Proposition 2.3.4. *Soient X et Y deux v.a. à valeurs $\mathbb{N}$ indépendantes. On a, pour tout $s \in [0,1]$,*

$$g_{X+Y}(s) = g_X(s)\, g_Y(s).$$

Preuve: On a, utilisant le th. 2.2.4,

$$g_{X+Y}(s) = \mathbb{E}(s^{X+Y}) = \mathbb{E}(s^X s^Y) = \mathbb{E}(s^X)\,\mathbb{E}(s^Y) = g_X(s)\, g_Y(s). \diamond$$

Exemples. (i) Soient X et Y deux v.a. indépendantes de loi $\mathcal{P}(\lambda)$ et $\mathcal{P}(\mu)$. On a

$$g_{X+Y}(s) = \mathrm{e}^{\lambda(s-1)}\mathrm{e}^{\mu(s-1)} = \mathrm{e}^{(\lambda+\mu)(s-1)}$$

et donc (prop. 2.3.2) $X + Y \sim \mathcal{P}(\lambda + \mu)$.

(ii) Soient $A_1, \ldots, A_n$ des événements indépendants de même probabilité $p = \mathbb{P}(A_k)$. Soient $S_n = 1_{A_1} + \ldots + 1_{A_n}$ le nombre d'événements réalisés, g la fonction génératrice (commune) de 1_{A_1} et g_n la fonction génératrice de S_n. On a $g(s) = \mathbb{E}(s 1_{A_1} + 1_{A_1^c}) = ps + 1 - p$. Donc $g_n(s) = [g(s)]^n = (ps + 1 - p)^n$ et (prop. 2.3.2) $S_n \sim B(n,p)$.

2.3.4. Critère d'indépendance. Soient X et Y deux v.a. à valeurs $\mathbb{N}$. On définit pour $u, v \in [0,1]$,

$$g_{(X,Y)}(u,v) = \sum_{m,n} \mathbb{P}(X = m, Y = n) u^m v^n = \mathbb{E}(u^X v^Y). \tag{2.13}$$

(Toujours avec la convention $0^0 = 1$). Alors $g_{(X,Y)}$ s'appelle la fonction génératrice du couple (X,Y).

Proposition 2.3.5. *Les v.a. à valeurs $\mathbb{N}$ X et Y sont indépendantes ssi, pour tous $u, v \in [0,1]$,*

$$g_{(X,Y)}(u,v) = g_X(u)\, g_Y(v). \tag{2.14}$$

Preuve: Si X et Y sont indépendantes, (2.14) résulte du th. 2.2.4. Réciproquement (2.14) s'écrit

$$\sum_{m,n} \mathbb{P}(X = m, Y = n) u^m v^m = \sum_m \mathbb{P}(X = m) u^m \sum_n \mathbb{P}(Y = n) v^n.$$

Appliquant $\frac{\partial^{m+n}}{\partial u^n \partial v^m}(0,0)$ aux deux membres, on obtient que, pour tous m, n,

$$\mathbb{P}(X = m, Y = n) = \mathbb{P}(X = m)\mathbb{P}(Y = n)$$

i.e. l'indépendance de X et Y. $\diamond$

La prop. 2.3.5 s'étend facilement au cas de n v.a.

Chapitre 3

Mesure. Intégration

Dans ce chapitre, on rappelle les résultats de la théorie de la mesure et de l'intégration qui seront utilisés par la suite.

3.1. Tribus

3.1.1. Soient E un ensemble et $\mathcal{B} \subset \mathcal{P}(E)$. On dit que $\mathcal{B}$ est une algèbre (resp. une tribu) si $E \in \mathcal{B}$, si $\mathcal{B}$ est stable par passage au complémentaire et par réunion et intersection finies (resp. dénombrables). Un couple $(E, \mathcal{B})$, $\mathcal{B}$ tribu sur E, s'appelle un espace mesurable. S'il est souvent possible de décrire les éléments d'une algèbre, il n'en est pas de même pour ceux d'une tribu. On remarque que $\mathcal{P}(E)$ est une tribu et que l'intersection d'une famille non vide quelconque de tribus est une tribu. Donc, étant donné $\mathcal{C} \subset \mathcal{P}(E)$, on peut considérer la plus petite tribu contenant $\mathcal{C}$, c'est l'intersection de toutes les tribus contenant $\mathcal{C}$. Cette tribu se note $\sigma(\mathcal{C})$ et s'appelle la tribu engendrée par $\mathcal{C}$. Le résultat suivant, appelé théorème de classe monotone, sera très utile par la suite.

Proposition 3.1.1. *Soient $\mathcal{C} \subset \mathcal{M} \subset \mathcal{P}(E)$. On suppose que $\mathcal{C}$ est stable par intersection finie, que $E \in \mathcal{M}$, que $A, B \in \mathcal{M}$ et $A \subset B$ impliquent $B \setminus A \in \mathcal{M}$ et que $\mathcal{M}$ est stable par limite croissante. Alors $\sigma(\mathcal{C}) \subset \mathcal{M}$.*

3.1.2. Supposons $E = \mathbb{R}^d$ et soit $\mathcal{O}$ la classe des ouverts de E. La tribu $\sigma(\mathcal{O})$ s'appelle la tribu borélienne de $\mathbb{R}^d$ et se note $\mathcal{B}(\mathbb{R}^d)$. Il est facile de voir qu'elle est aussi engendrée par les fermés, par les boules, par les pavés et même par les pavés à coordonnées rationnelles (cette dernière famille ayant l'avantage d'être dénombrable). Si $d = 1$, on considérera, outre $\mathcal{B}(\mathbb{R})$, $\mathcal{B}(\mathbb{R}^+) = \{A \in \mathcal{B}(\mathbb{R}), A \subset \mathbb{R}^+\}$, $\mathcal{B}(\overline{\mathbb{R}}) = \sigma(\mathcal{B}(\mathbb{R}), \{+\infty\}, \{-\infty\})$ et $\mathcal{B}(\overline{\mathbb{R}^+}) = \sigma(\mathcal{B}(\mathbb{R}^+), \{+\infty\})$. On étend les opérations usuelles à $\overline{\mathbb{R}^+}$ en posant $(+\infty) \times 0 = 0 \times (+\infty) = 0$.

3.1.3. Soient $(E_1, \mathcal{B}_1)$ et $(E_2, \mathcal{B}_2)$ deux espaces mesurables. Une application de E_1 dans E_2 est dite mesurable si, pour tout $A \in \mathcal{B}_2$, $f^{-1}(A) \in \mathcal{B}_1$. Il est facile de voir que, pour cela, il suffit que $f^{-1}(A) \in \mathcal{B}_1$ pour tout $A \in \mathcal{C}$ avec $\sigma(\mathcal{C}) = \mathcal{B}_2$. Ceci

implique que, si f est continue de $\mathbb{R}^d$ dans $\mathbb{R}^m$, f est borélienne i.e. mesurable pour les tribus boréliennes. De plus, cette notion est transitive i.e. la composée de deux applications mesurables est mesurable. Quand l'espace d'arrivée est $\mathbb{R}$, $\overline{\mathbb{R}}$, $\overline{\mathbb{R}^+}$, $\mathbb{R}^d$, $\mathbb{C}$, il est toujours supposé muni de sa tribu borélienne.

3.1.4. Soit $(E, \mathcal{B})$ un espace mesurable. Pour qu'une application numérique soit mesurable, il suffit que, pour tout $a \in \mathbb{R}$, $\{f > a\} := \{x,\ f(x) > a\} \in \mathcal{B}$. On peut aussi considérer $\{f < a\}$, $\{f \leq a\}$, $\{f \geq a\}$. Ceci implique que, si f, g, f_n sont des fonctions numériques mesurables, il en est de même de $-f$, $\sup(f,g)$, $\inf(f,g)$, $f^+ = \sup(f,0)$, $f^- = \sup(-f,0)$, $\sup f_n$, $\inf f_n$, $\limsup f_n$, $\liminf f_n$, $\lim f_n$ si elle existe.

Rappelons que, notant $f_n \uparrow f$ (resp.$f_n \downarrow f$) si, pour tout $x \in E$, $f_n(x)$ croît (resp. décroît) vers $f(x)$,

$$\limsup f_n(x) = \lim_n \downarrow \sup_{k \geq n} f_k(x), \quad \liminf f_n(x) = \lim_n \uparrow \inf_{k \geq n} f_k(x), \tag{3.1}$$

ces quantités étant à valeurs $\overline{\mathbb{R}}$ et que $f = \lim f_n$ ssi $\limsup f_n = \liminf f_n = f$.

Soient f, g des fonctions numériques mesurables. Alors $\phi : x \mapsto (f(x), g(x))$ est mesurable de $(E, \mathcal{B})$ dans $\mathbb{R}^2$ puisque $\phi^{-1}(A \times B) = f^{-1}(A) \cap g^{-1}(B)$. Ceci implique que, si H est une application borélienne de $\mathbb{R}^2$ dans $\mathbb{R}$, $H(f,g)$ est mesurable. On en déduit que $f+g$, fg, $\frac{f}{g}$, si elle existe, sont mesurables.

3.1.5. Pour $A \subset B$, on appelle fonction indicatrice de A et on note 1_A la fonction valant 1 sur A et 0 sur A^c (on note A^c le complémentaire de A). On a

$$1_{A^c} = 1 - 1_A, \quad 1_{\cap A_n} = \prod_n 1_{A_n} = \inf 1_{A_n}, \quad 1_{\cup A_n} = \sup 1_{A_n}.$$

Une application f de E muni de la tribu $\mathcal{B}$ dans $\mathbb{R}$ est dite étagée si elle s'écrit $f = \sum_{k=1}^n a_k 1_{A_k}$, $A_k \in \mathcal{B}$. On notera
$[\mathcal{B}]$ l'ensemble des fonctions réelles $\mathcal{B}$-mesurables,
$b\mathcal{B}$ l'ensemble des fonctions réelles $\mathcal{B}$-mesurables bornées,
$\mathcal{B}^+$ l'ensemble des fonctions $\mathcal{B}$-mesurables à valeurs $\overline{\mathbb{R}^+}$,
$e\mathcal{B}^+$ l'ensemble des fonctions étagées positives.

Le résultat suivant est à la base de la construction de l'intégrale

Proposition 3.1.2. *Toute $f \in \mathcal{B}^+$ est limite d'une suite croissante de fonctions de $e\mathcal{B}^+$.*

Preuve: Il suffit de considérer

$$f_n(x) = \sum_{k=0}^{n2^n - 1} \frac{k}{2^n} 1_{\{\frac{k}{2^n} \leq f(x) < \frac{k+1}{2^n}\}} + n 1_{\{f(x) \geq n\}}. \tag{3.2}$$

3.1.6. Soit f une application de E dans un espace mesurable $(A, \mathcal{A})$. On note $\sigma(f)$ et on appelle tribu engendrée par f la plus petite tribu sur E rendant f mesurable. On a donc $\sigma(f) = \{f^{-1}(A),\ A \in \mathcal{A}\}$.

Proposition 3.1.3. *Soient $f : E \to (A, \mathcal{A})$ et $h : E \to \mathbb{R}$ (resp. $E \to \overline{\mathbb{R}^+}$). Alors h est $\sigma(f)$-mesurable ssi il existe $g \in [\mathcal{A}]$ (resp. $g \in \mathcal{A}^+$) telle que $h = g \circ f$.*

Preuve: Evidemment si $h = g \circ f$, h est $\sigma(f)$-mesurable (transitivité). Réciproquement supposons d'abord $h \in e[\sigma(f)]^+$, on a $h = \sum_{k=1}^n a_k 1_{B_k}$ avec $B_k \in \sigma(f)$ et donc $B_k = f^{-1}(A_k)$, $A_k \in \mathcal{A}$. Vu que $1_{B_k} = 1_{A_k} \circ f$, on a $h = g \circ f$ avec $g = \sum_{k=1}^n a_k 1_{A_k}$. Si $h \in [\sigma(f)]^+$, on a $h = \lim \uparrow h_n$ avec $h_n \in e[\sigma(f)]^+$ et donc $h_n = g_n \circ f$, $g_n \in \mathcal{A}^+$. On en déduit $h = g \circ f$ avec $g = \limsup g_n \in \mathcal{A}^+$. Si $h \in [\sigma(f)]$, on a $h = h^+ - h^-$ et $h^+ = g_1 \circ f$, $h^- = g_2 \circ f$ avec $g_i \in \mathcal{A}^+$. On a alors $h = g \circ f$ avec $g = g_1 1_{\{g_1 < +\infty\}} - g_2 1_{\{g_2 < +\infty\}} \in [\mathcal{A}]$. $\diamond$

Plus généralement si $(f_i,\, i \in I)$ est une famille d'applications de E dans des espaces mesurables $(F_i, \mathcal{F}_i)$, on note $\sigma(f_i,\, i \in I)$ et on appelle tribu engendrée par les f_i la plus petite tribu sur E rendant toutes les f_i mesurables. On a donc

$$\sigma(f_i,\, i \in I) = \sigma(f_i^{-1}(A_i),\, A_i \in \mathcal{F}_i,\, i \in I).$$

3.2. Mesures

3.2.1. Soit $(E, \mathcal{B})$ un espace mesurable.

Définition 3.2.1. *On appelle mesure sur $(E, \mathcal{B})$ toute application μ de $\mathcal{B}$ dans $\overline{\mathbb{R}^+}$ telle que*
(i) $\mu(\emptyset) = 0$,
(ii) pour tous $A_n \in \mathcal{B}$ deux à deux disjoints, $\mu(\cup_n A_n) = \sum_n \mu(A_n)$.
Le triplet $(E, \mathcal{B}, \mu)$ s'appelle un espace mesuré.

Propriétés: (i) si $A, B \in \mathcal{B}$ et $A \subset B$, $\mu(A) \le \mu(B)$,

(ii) si $A_n \in \mathcal{B}$, $\mu(\cup_n A_n) \le \sum_n \mu(A_n)$,

(iii) si $A_n \in \mathcal{B}$ et si $A_n \uparrow A$ (i.e. $1_{A_n} \uparrow 1_A$), $\mu(A_n) \uparrow \mu(A)$,

(iv) si $A_n \in \mathcal{B}$, si $A_n \downarrow A$ (i.e. $1_{A_n} \downarrow 1_A$) et si, pour un n_0, $\mu(A_{n_0}) < +\infty$, $\mu(A_n) \downarrow \mu(A)$.

Si $E = \cup_n E_n$ avec $E_n \in \mathcal{B}$ et $\mu(E_n) < +\infty$, la mesure μ est dite σ-finie. Si $\mu(E) < +\infty$, la mesure μ est dite bornée. Si $\mu(E) = 1$, la mesure μ est appelée une probabilité.

Exemple. Soit $a \in E$. alors $\delta_a(A) = 1_A(a)$ définit une mesure sur $(E, \mathcal{B})$ appelée mesure de Dirac de a. Plus géralement, étant donnés $a_n \in E$ et $\lambda_n \ge 0$, $\mu = \sum_n \lambda_n \delta_{a_n}$ est une mesure sur $(E, \mathcal{B})$ (prop. 2.1.2).

Remarque. La propriété (ii) de la def. 3.2.1 s'appelle σ-additivité. Si dans la def. 3.2.1, on suppose que $\mathcal{B}$ est seulement une algèbre, la définition a encore un sens en rajoutant dans (ii) la condition $\cup_n A_n \in \mathcal{B}$. On a ainsi la notion de mesure sur une algèbre.

Proposition 3.2.2. *Soient μ et ν deux mesures sur $(E, \mathcal{B})$ et $\mathcal{C} \subset \mathcal{B}$ une classe d'ensembles stable par intersection finie. On suppose que, pour tout $A \in \mathcal{C}$, $\mu(A) = \nu(A) < +\infty$ et que $E = \lim \uparrow E_n$ avec $E_n \in \mathcal{C}$. Alors $\mu(A) = \nu(A)$ pour tout $A \in \sigma(\mathcal{C})$.*

Preuve: Supposons d'abord $\mu(E) = \nu(E) < +\infty$. Soit $\mathcal{M} = \{A \in \mathcal{B},\ \mu(A) = \nu(A)\}$. On vérifie immédiatement que les hypothèses de la prop. 3.1.2 sont vérifiées. On a donc $\sigma(\mathcal{C}) \subset \mathcal{M}$. Le cas général se traite en appliquant ce résultat aux mesures $\mu_n(A) = \mu(A \cap E_n)$ et $\nu_n(A) = \nu(A \cap E_n)$. $\diamond$

Corollaire 3.2.3. *Soient μ et ν deux probabilités sur $(E, \mathcal{B})$ et $\mathcal{C} \subset \mathcal{B}$ une classe d'ensembles stable par intersection finie telle que $\sigma(\mathcal{C}) = \mathcal{B}$. Si $\mu(A) = \nu(A)$ pour tout $A \in \mathcal{C}$, alors $\mu = \nu$.*

3.2.2. Soit $(E, \mathcal{B}, \mu)$ un espace mesuré. Un sous-ensemble A de E est dit négligeable (ou μ-négligeable s'il y a ambiguïté) si $A \subset B$ avec $B \in \mathcal{B}$ et $\mu(B) = 0$. Une propriété est vraie presque partout (en abrégé p.p. ou, plus présisemment, μ p.p.) si elle est vraie en dehors d'un ensemble négligeable. Par exemple $f = g$ p.p. signifie que $\{x \in E,\ f(x) \neq g(x)\}$ est négligeable. Si μ est une probabilité, on dit presque sûrement (en abrégé p.s.) pour presque partout. On note $\mathcal{N}$ la classe des ensembles négligeables. Il faut noter que si $A_n \in \mathcal{N}$, on a $\cup_n A_n \in \mathcal{N}$. Si $\mathcal{N} \subset \mathcal{B}$, l'espace mesuré $(E, \mathcal{B}, \mu)$ est dit complet. Si ce n'est pas le cas, on peut le "compléter" de la façon suivante. On définit $\overline{\mathcal{B}} = \sigma(\mathcal{B}, \mathcal{N})$. Alors $A \in \overline{\mathcal{B}}$ ssi $A = B \cup N$ avec $B \in \mathcal{B}$ et $N \in \mathcal{N}$. On peut prolonger μ à $\overline{\mathcal{B}}$ en posant $\mu(A) = \mu(B)$ (il est facile de voir que ceci ne dépend pas de l'écriture de A). L'espace $(E, \overline{\mathcal{B}}, \mu)$ est alors complet et s'appelle le complété de $(E, \mathcal{B}, \mu)$. Enfin on vérifie aisément que $f : E \to \overline{\mathbb{R}}$ est $\overline{\mathcal{B}}$−mesurable ssi il existe $g, h : E \to \overline{\mathbb{R}}$ $\mathcal{B}$−mesurables telles que $g \leq f \leq h$ et $g = h$ μ p.p.

3.2.3. Construction. Dans la suite, la plupart du temps, on partira d'un espace mesurable ou d'un espace de probabilité sans se soucier de sa construction. Il est néanmoins indispensable de s'assurer de l'existence de tels objets. On va s'intéresser aux mesures sur $\mathcal{B}(\mathbb{R})$ finies sur les intervalles bornés. Observons d'abord que $\mathcal{C} = \{\,]a, b],\ -\infty < a < b < +\infty\}$ est une classe stable par intersection finie et que $\sigma(\mathcal{C}) = \mathcal{B}(\mathbb{R})$. Il résulte alors de la prop. 3.2.2 qu'une mesure μ sur $\mathcal{B}(\mathbb{R})$ finie sur les intervalles bornés est déterminée par les valeurs $\mu(]a, b])$. Ensuite, étant donnée une telle mesure, si on pose

$$F(0) = 0;\ \ F(x) = \mu(]0, x]),\ x > 0;\ \ F(x) = -\mu(]x, 0]),\ x < 0,$$

$F(x)$ est une fonction continue à droite et croissante et l'on a $\mu(]a, b]) = F(b) - F(a)$. On est donc ramené au problème suivant. Soit F une application de $\mathbb{R}$ dans $\mathbb{R}$ continue à droite et croissante, existe-t-il une mesure μ sur $\mathcal{B}(\mathbb{R})$ telle que $\mu(]a, b]) = F(b) - F(a)$? Il est facile de décrire l'algèbre $\mathcal{A}$ engendrée par $\mathcal{C}$, on a

$$\mathcal{A} = \{\, A = \cup_{k=1}^{n}]a_k, b_k],\ -\infty \leq a_1 < b_1 < a_2 < \ldots < b_{n-1} < a_n < b_n \leq +\infty\}$$

en convenant que, si $b_n = +\infty$, $]a_n, b_n] =]a_n, +\infty[$. On définit μ sur $\mathcal{A}$ par $\mu(A) = \sum_{k=1}^n F(b_k) - F(a_k)$ où $F(+\infty) = \lim_{x\to+\infty} F(x)$, $F(-\infty) = \lim_{x\to-\infty} F(x)$. Il est facile de montrer que μ est additive sur $\mathcal{A}$, un peu plus délicat de montrer que μ est σ-additive sur $\mathcal{A}$ mais cela se fait. On a donc construit une mesure μ sur $\mathcal{A}$ telle que $\mu(]a, b]) = F(b) - F(a)$. Pour passer à $\mathcal{B}(\mathbb{R})$, on utilise le théorème de Carathéodory:

Théorème 3.2.4. *Soit μ une mesure sur une algèbre $\mathcal{A}$, alors μ se prolonge en une mesure sur $\sigma(\mathcal{A})$. De plus, si μ est σ-finie, ce prolongement est unique.*

Tout ceci donne, puisque dans notre cas $\sigma(\mathcal{A}) = \mathcal{B}(\mathbb{R})$,

Théorème 3.2.5. *Soit F une application de $\mathbb{R}$ dans $\mathbb{R}$ continue à droite et croissante. Il existe une et une seule mesure μ sur $\mathcal{B}(\mathbb{R})$ telle que, pour tous $a < b$, $\mu(]a, b]) = F(b) - F(a)$.*

Si on choisit $F(x) = x$, on obtient l'existence et l'unicité d'une mesure λ sur $\mathcal{B}(\mathbb{R})$ vérifiant, pour tout intervalle I, $\lambda(I) = |I|$. C'est la mesure de Lebesgue sur $\mathbb{R}$. Si $\mathcal{N}$ est la classe des ensembles λ-négligeables, $\overline{\mathcal{B}(\mathbb{R})} = \sigma(\mathcal{B}, \mathcal{N})$ s'appelle la tribu des ensembles Lebesgue-mesurables (elle est beaucoup plus "grosse" que $\mathcal{B}(\mathbb{R})$) et λ se prolonge sans peine à $\overline{\mathcal{B}(\mathbb{R})}$ comme en 3.2.2.

3.3. Intégration

Soit $(E, \mathcal{B}, \mu)$ un espace mesuré.

3.3.1. Intégration des fonctions positives. On va construire l'intégrale de f par rapport à μ. Si $f \in e\mathcal{B}^+$, c'est très facile, f s'écrit $f = \sum_{k=1}^n a_k 1_{A_k}$, $A_k \in \mathcal{B}$ et l'on pose

$$\int f\, d\mu := \sum_{k=1}^n a_k \mu(A_k).$$

Des considérations élémentaires montrent que ceci ne dépend pas de l'écriture de f et que, pour $f, g \in e\mathcal{B}^+$ et $a, b \in \mathbb{R}^+$, $\int (af + bg)\, d\mu = a \int f\, d\mu + b \int g\, d\mu$ et que, si $f \le g$, $\int f\, d\mu \le \int g\, d\mu$. On a aussi le résultat plus technique suivant qui est la clé de la construction.

Lemme 3.3.1. *Si $f_n, g_n \in e\mathcal{B}^+$ sont croissantes et si $\lim \uparrow f_n = \lim \uparrow g_n$, on a $\lim \uparrow \int f_n\, d\mu = \lim \uparrow \int g_n\, d\mu$.*

Soit $f \in \mathcal{B}^+$. Il existe (prop. 3.1.2) une suite $f_n \in e\mathcal{B}^+$ telle que $f_n \uparrow f$, on a alors $\int f_n\, d\mu \uparrow$ et on pose $\int f\, d\mu = \lim \uparrow \int f_n\, d\mu$. Le point important est que, d'après le lem. 3.3.1, cette limite ne dépend pas de la suite f_n choisie. On a en particulier, vu (3.2), pour $f \in \mathcal{B}^+$,

$$\int f\, d\mu = \lim \uparrow \sum_{k=0}^{n2^n-1} \frac{k}{2^n} \mu(\{x,\ \frac{k}{2^n} \le f(x) < \frac{k+1}{2^n}\}) + n\mu(\{x,\ f(x) \ge n\}). \qquad (3.3)$$

Par passage à la limite, on obtient immédiatement que, pour $f, g \in \mathcal{B}^+$ et $a, b \in \mathbb{R}^+$, $\int (af + bg)\, d\mu = a \int f\, d\mu + b \int g\, d\mu$ et que, si $f \leq g$, $\int f\, d\mu \leq \int g\, d\mu$. Enfin on dira que $f \in \mathcal{B}^+$ est intégrable si $\int f\, d\mu < +\infty$.

3.3.2. Intégration des fonctions réelles ou complexes. On pose

$$\mathcal{L}^1 = \mathcal{L}^1(E, \mathcal{B}, \mu) = \{f \in [\mathcal{B}],\ \int |f|\, d\mu < +\infty\}. \tag{3.4}$$

Si $f \in \mathcal{L}^1$, f^+ et f^- sont intégrables et on pose

$$\int f\, d\mu = \int f^+\, d\mu - \int f^-\, d\mu.$$

Il est facile de voir (vu que $|f + g| \leq |f| + |g|$) que $\mathcal{L}^1$ est un espace vectoriel et que $f \mapsto \int f\, d\mu$ est une forme linéaire positive sur $\mathcal{L}^1$. De plus, pour $f \in \mathcal{L}^1$, $|\int f\, d\mu| \leq \int |f|\, d\mu$.

Si f est $\mathcal{B}$-mesurable à valeurs $\mathbb{C}$, on pose ($|f|$ désignant le module),

$$\mathcal{L}^1_{\mathbb{C}} = \mathcal{L}^1_{\mathbb{C}}(E, \mathcal{B}, \mu) = \{f\ \mathcal{B}\text{-mesurable complexe},\ \int |f|\, d\mu < +\infty\}. \tag{3.5}$$

On définit alors, pour $f \in \mathcal{L}^1_{\mathbb{C}}$, $\int f\, d\mu = \int \Re(f)\, d\mu + i \int \Im(f)\, d\mu$. $\mathcal{L}^1_{\mathbb{C}}$ est un espace vectoriel sur $\mathbb{C}$ et $f \mapsto \int f\, d\mu$ une forme linéaire sur $\mathcal{L}^1_{\mathbb{C}}$. On a aussi, pour $f \in \mathcal{L}^1_{\mathbb{C}}$, $|\int f\, d\mu| \leq \int |f|\, d\mu$.

3.3.3. Propriétés.

(i) Si $f \in \mathcal{B}^+$ et si $\int f\, d\mu < +\infty$, $f < +\infty$ p.p.

(ii) Si $f \in \mathcal{B}^+$ et si $\int f\, d\mu = 0$, $f = 0$ p.p.

(iii) Si $f, g \in \mathcal{L}^1$ et si $f \leq g$ p.p., $\int f\, d\mu \leq \int g\, d\mu$.

(iv) Si $f \in \mathcal{L}^1_{\mathbb{C}}$ et si $A \in \mathcal{B}$, $f 1_A \in \mathcal{L}^1_{\mathbb{C}}$. On pose alors

$$\int_A f\, d\mu := \int f 1_A\, d\mu,\ \ A \in \mathcal{B},\ f \in \mathcal{L}^1_{\mathbb{C}} \cup \mathcal{B}^+.$$

(v) Si $f \in \mathcal{L}^1$ et si, pour tout $A \in \mathcal{B}$, $\int_A f\, d\mu \geq 0$ alors $f \geq 0$ p.p.

(vi) Si $f, g \in \mathcal{L}^1$ et si, pour tout $A \in \mathcal{B}$, $\int_A f\, d\mu \leq \int_A g\, d\mu$, alors $f \leq g$ p.p.

Il nous reste à énoncer les résultats concernant les passages à la limite. Le premier d'où découlent facilement les autres s'appelle théorème de convergence monotone ou théorème de Beppo-Levi.

Théorème 3.3.2. *Soit $f_n \in \mathcal{B}^+$ une suite croissante, alors*

$$\lim \uparrow \int f_n\, d\mu = \int \lim \uparrow f_n\, d\mu.$$

Corollaire 3.3.3. *Soit* $g_n \in \mathcal{B}^+$, *alors*

$$\sum_n \int g_n \, d\mu = \int \sum_n g_n \, d\mu.$$

Proposition 3.3.4. *(Lemme de Fatou) (i) Soit* $f_n \in \mathcal{B}^+$, *alors*

$$\int \liminf f_n \, d\mu \leq \liminf \int f_n \, d\mu.$$

(ii) Soit $f_n \in [\mathcal{B}]$ *avec* $|f_n| \leq g \in \mathcal{L}^1$, *alors*

$$\int \liminf f_n \, d\mu \leq \liminf \int f_n \, d\mu \leq \limsup \int f_n \, d\mu \leq \int \limsup f_n \, d\mu.$$

(ii) implique le célèbre théorème de Lebesgue,

Théorème 3.3.5. *Soit* $f_n \in \mathcal{L}^1_{\mathbb{C}}$ *telles que* $f_n \to f$ *p.p. avec* $|f_n| \leq g \in \mathcal{L}^1$, *alors*

$$\lim \int f_n \, d\mu = \int f \, d\mu.$$

Ce théorème a une version "continue" très utile.

Corollaire 3.3.6. *Soit* $(f_t,\, t \in U)$ *une famille d'éléments de* $\mathcal{L}^1_{\mathbb{C}}$, U *ouvert de* $\mathbb{R}^d$. *On suppose que* $\lim_{t \to t_0} f_t = f$ *p.p. et que, pour tout* $t \in U$, $|f_t| \leq g \in \mathcal{L}^1$, *alors* $\lim_{t \to t_0} \int f_t \, d\mu = \int f \, d\mu$.

Preuve: Il suffit de remarquer que $\lim_{t \to t_0} \int f_t \, d\mu = \int f \, d\mu$ ssi, pour toute suite t_n tendant vers t_0, $\lim_{t_n \to t_0} \int f_{t_n} \, d\mu = \int f \, d\mu$ et d'appliquer le th. 3.3.5. $\diamond$

Donnons un exemple d'utilisation de ce corollaire.

Proposition 3.3.7. *Soient* $(E, \mathcal{B}, \mu)$ *un espace mesuré,* I *un intervalle ouvert et* $(f(t,x),\, t \in I)$ *une famille d'éléments de* $\mathcal{L}^1_{\mathbb{C}}(\mu)$. *On pose, pour tout* $t \in I$, $\phi(t) = \int f(t,x) \, d\mu(x)$. *On suppose que, pour tout* $x \in A$, $t \mapsto f(t,x)$ *est dérivable sur* I, *que, pour tous* $x \in A$ *et* $t \in I$, $|\frac{\partial f}{\partial t}(t,x)| \leq g(x)$, *que* $g \in \mathcal{L}^1(\mu)$ *et que* $\mu(A^c) = 0$. *Alors* ϕ *est dérivable sur* I *et* $\phi'(t) = \int \frac{\partial f}{\partial t}(t,x) \, d\mu(x)$.

Preuve: On a

$$\frac{1}{h}(\phi(t+h) - \phi(t)) = \int_A \frac{1}{h}(f(t+h,x) - f(t,x)) \, d\mu(x).$$

D'après la formule des accroissements finis, on a, pour $x \in A$,

$$|\frac{1}{h}(f(t+h,x) - f(t,x))| = |\frac{\partial f}{\partial t}(\theta, x)| \leq g(x)$$

si h est assez petit et

$$\frac{1}{h}(f(t+h,x) - f(t,x)) \to_{h \to 0} \frac{\partial f}{\partial t}(t,x).$$

On peut appliquer le cor. 3.3.6 et

$$\int_A \frac{1}{h}(f(t+h,x)-f(t,x))\,d\mu(x) \to_{h\to 0} \int_A \frac{\partial f}{\partial t}(t,x)\,d\mu(x) = \int \frac{\partial f}{\partial t}(t,x)\,d\mu(x). \diamond$$

3.3.4. Lien avec l'intégrale usuelle. Soit f une fonction réelle continue sur $[a,b]$ et posons, pour $a \le x \le b$, $F(x) = \int_a^x f(t)\,dt$ (intégrale au sens usuelle) et $G(x) = \int 1_{[a,a+x[} f\,d\lambda$, λ mesure de Lebesgue sur $\mathbb{R}$. On sait que $F(a) = 0$, F est continue sur $[a,b]$ et que, sur $]a,b[$, F est dérivable avec $F' = f$. Il est facile de vérifier que G a les mêmes propriétés. Ceci implique que $F = G$ sur $[a,b]$ et, en particulier, que

$$\int_a^b f(t)\,dt = \int 1_{[a,b[} f\,d\lambda.$$

Par additivité, cette formule est encore vraie si f est continue par morceaux sur $[a,b]$.

Considérons maintenant une application f de $\mathbb{R}$ dans $\mathbb{R}$ continue par morceaux telle que $\int_{-\infty}^{+\infty} f(t)\,dt$ soit absolument convergente. Lorsque $a \downarrow -\infty$ et $b \uparrow +\infty$, d'une part, par définition, $\int_a^b |f(t)|\,dt \to \int_{-\infty}^{+\infty} |f(t)|\,dt < +\infty$ et $\int_a^b f(t)\,dt \to \int_{-\infty}^{+\infty} f(t)\,dt$; d'autre part, $\int 1_{[a,b[}|f|\,d\lambda \to \int |f|\,d\lambda$ (convergence monotone) ce qui implique que $f \in \mathcal{L}^1(\lambda)$ puis $\int 1_{[a,b[} f\,d\lambda \to \int f\,d\lambda$ (théorème de Lebesgue puisque $|1_{[a,b[}f| \le |f| \in \mathcal{L}^1(\lambda)$). Donc

$$\int_{-\infty}^{+\infty} f(t)\,dt = \int f\,d\lambda.$$

Par contre, si $\int_{-\infty}^{+\infty} f(t)\,dt$ est convergente mais pas absolument convergente (par exemple $f(x) = \frac{\sin x}{x}$), $f \notin \mathcal{L}^1(\lambda)$.

3.3.5. Espaces L^p. Soit $(E,\mathcal{B},\mu)$ un espace mesuré. On note $\mathcal{L}^0$ l'ensemble des applications $\mathcal{B}$-mesurables de E dans $\overline{\mathbb{R}}$ finies p.p. On dit que $f \sim g$ si $f = g$ p.p. Alors $\sim$ est une relation d'équivalence sur $\mathcal{L}^0$. On note $L^0 = \mathcal{L}^0/\sim$. En fait L^0 est l'espace des classes de fonctions $\mathcal{B}$-mesurables définies à un p.p. près. Puisque $f = g$ p.p. implique $\int |f|\,d\mu = \int |g|\,d\mu$ et $\int f\,d\mu = \int g\,d\mu$ si f et g sont dans $\mathcal{L}^1$, on peut définir sans ambiguïté, pour $f \in L^0$, $\int |f|\,d\mu$ puis, si $\int |f|\,d\mu < +\infty$, $\int f\,d\mu$. Par abus de langage, *dans toute la suite nous noterons de la même façon une fonction et sa classe d'équivalence.* On pose alors, pour $1 \le p < +\infty$ et $f \in L^0$,

$$||f||_p = [\int |f|^p\,d\mu]^{\frac{1}{p}}$$

et, pour $p = +\infty$,

$$||f||_\infty = \inf(M,\ \mu(|f| > M) = 0).$$

On a deux inégalités fondamentales. Pour $f, g \in L^0_+$,

$$||f+g||_p \le ||f||_p + ||g||_p,\ \ 1 \le p \le +\infty \tag{3.6}$$

qui s'appelle l'inégalité de Minkowski et

$$||fg||_1 \le ||f||_p\,||g||_q,\ \ 1 \le p \le +\infty,\ \frac{1}{p}+\frac{1}{q} = 1 \tag{3.7}$$

qui s'appelle l'inégalité de Hölder. Notons que pour $p = q = 2$, (3.7) implique l'inégalité de Schwarz

$$[\int |fg|\,d\mu]^2 \leq (\int f^2\,d\mu)(\int g^2\,d\mu).$$

On note

$$\mathcal{L}^p = \{f \in \mathcal{L}^0,\ \int |f|^p\,d\mu < +\infty\}, \quad L^p = \{f \in L^0,\ \int |f|^p\,d\mu < +\infty\}.$$

Alors L^p muni de la norme $||.||_p$ est un espace de Banach et L^2 est un espace de Hilbert pour le produit scalaire

$$< f, g >= \int fg\,d\mu.$$

On peut aussi considérer le cas des fonctions à valeurs complexes. On définit de la même façon $L^p_{\mathbb{C}} = L^p_{\mathbb{C}}(E, \mathcal{B}, \mu)$. Il faut noter que $L^2_{\mathbb{C}}$ est associé au produit scalaire

$$< f, g >= \int f\bar{g}\,d\mu.$$

Proposition 3.3.8. *Pour* $1 \leq p < +\infty$, $\mathcal{E}^0 = \{f,\ f = \sum_{k=1}^n a_k 1_{A_k},\ A_k \in \mathcal{B},\ \mu(A_k) < +\infty\}$ *est dense dans* $L^p(E, \mathcal{B}, \mu)$.

Preuve: Il suffit de considérer $f \geq 0$. Alors il existe (prop. 3.1.2) une suite $f_n \in e\mathcal{B}^+$ telle que $f_n \uparrow f$. Vu que $f_n^p \leq f^p \in \mathcal{L}^1$, $f_n \in \mathcal{E}^0$. On a, puisque $f < +\infty$ p.p., $|f - f_n|^p \to 0$ p.p. et $|f - f_n|^p \leq f^p \in \mathcal{L}^1$ donc (th. de Lebesgue) $\int |f - f_n|^p\,d\mu \to 0$. $\diamond$

3.4. Mesures à densité

3.4.1. Soit μ une mesure sur $(E, \mathcal{B})$. On peut lui associer une application I de $\mathcal{B}^+$ dans $\overline{\mathbb{R}^+}$ en posant $I(f) = \int f\,d\mu$, $f \in \mathcal{B}^+$. L'application I a les propriétés suivantes: $I(f + g) = I(f) + I(g)$, $I(af) = aI(f)$, $a \in \mathbb{R}^+$ et $I(f_n) \uparrow I(f)$ si $f_n \uparrow f$. Réciproquement on a,

Proposition 3.4.1. *Soient* $(E, \mathcal{B})$ *un espace mesurable et* I *une application de* $\mathcal{B}^+$ *dans* $\overline{\mathbb{R}^+}$ *telle que*
(i) si $f, g \in \mathcal{B}^+$, $I(f + g) = I(f) + I(g)$; *si* $f \in \mathcal{B}^+$ *et* $a \in \mathbb{R}^+$, $I(af) = aI(f)$,
(ii) si $f_n \in \mathcal{B}^+$ *et si* $f_n \uparrow f$, $I(f_n) \uparrow I(f)$.
Alors $\mu(A) = I(1_A)$, $A \in \mathcal{B}$, *définit une mesure sur* $\mathcal{B}$ *et on a, pour toute* $f \in \mathcal{B}^+$, $I(f) = \int f\,d\mu$.

Preuve: Soient $A_n \in \mathcal{B}$ des ensembles deux à deux disjoints d'union A, on a $1_A = \sum_n 1_{A_n} = \lim \uparrow \sum_{k=1}^n 1_{A_k}$ et

$$\mu(A) = I(1_A) = I(\lim \uparrow \sum_{k=1}^n 1_{A_k}) = \lim \uparrow I(\sum_{k=1}^n 1_{A_k}) = \lim \uparrow \sum_{k=1}^n I(1_{A_k}) = \sum_n \mu(A_n).$$

Ce qui montre que μ est une mesure. On a alors, pour toute $f \in e\mathcal{B}^+$, $I(f) = \int f\,d\mu$. On conclut facilement en utilisant la prop. 3.1.2. $\diamond$

3.4.2. Mesures à densité.

Proposition 3.4.2. *Soient $(E, \mathcal{B}, \mu)$ un espace mesuré et $h \in \mathcal{B}^+$. La formule $\nu(A) = \int_A h\,d\mu$, $A \in \mathcal{B}$ définit une mesure sur $\mathcal{B}$ appelée mesure de densité h par rapport à μ et notée $h.\mu$. On a, pour toute $f \in \mathcal{B}^+$,*

$$\int f\,d\nu = \int fh\,d\mu. \tag{3.8}$$

De plus $f \in [\mathcal{B}]$ est ν-intégrable ssi fh est μ-intégrable et l'on a dans ce cas (3.8).

Preuve: On considère la fonctionnelle $I(f) = \int fh\,d\mu$, $f \in \mathcal{B}^+$ et on applique la prop. 3.4.1. La dernière assertion est pure routine en écrivant $f = f^+ - f^-$. $\diamond$

Supposons que $\nu = h_1.\mu = h_2.\mu$ et que ν soit bornée, alors $h_1, h_2 \in \mathcal{L}^1(\mu)$ et on a (3.3.3 (vi)) $h_1 = h_2$ μ p.p. On voit facilement que ceci est encore vrai si ν est σ-finie.

3.4.3. Théorème de Radon-Nikodym. Soient μ, ν deux mesures sur $(E, \mathcal{B})$. On cherche à savoir si ν a une densité par rapport à μ. Si $\nu = h.\mu$, on a évidemment, pour $A \in \mathcal{B}$, $\mu(A) = 0$ implique $\nu(A) = 0$. Il est remarquable que cette propriété suffise à caractériser les mesures ayant une densité par rapport à μ.

Définition 3.4.3. *On dit que ν est absolument continue par rapport à μ si*

$$A \in \mathcal{B} \text{ et } \mu(A) = 0 \text{ impliquent } \nu(A) = 0.$$

On note alors $\nu \ll \mu$. On a (théorème de Radon-Nikodym):

Théorème 3.4.4. *Soient μ, ν deux mesures σ-finies sur $(E, \mathcal{B})$ telles que $\nu \ll \mu$. Alors il existe $h \in \mathcal{B}^+$, unique à un μ p.p. près, telle que $\nu = h.\mu$.*

3.5. Mesures produits

3.5.1. Soient $(E_1, \mathcal{B}_1)$ $(E_2, \mathcal{B}_2)$ deux espaces mesurables. On définit une tribu sur $E_1 \times E_2$, appelée tribu produit de $\mathcal{B}_1$ et $\mathcal{B}_2$ et notée $\mathcal{B}_1 \otimes \mathcal{B}_2$, par

$$\mathcal{B}_1 \otimes \mathcal{B}_2 = \sigma(A_1 \times A_2,\ A_1 \in \mathcal{B}_1,\ A_2 \in \mathcal{B}_2).$$

Alors si $f : E_1 \times E_2 \to \overline{\mathbb{R}^+}$ est une fonction $\mathcal{B}_1 \otimes \mathcal{B}_2$-mesurable, on a que pour tout $x_1 \in E_1$, $x_2 \mapsto f(x_1, x_2)$ est $\mathcal{B}_2$-mesurable et que, pour tout $x_2 \in E_2$, $x_1 \mapsto f(x_1, x_2)$ est $\mathcal{B}_1$-mesurable. En particulier si $A \in \mathcal{B}_1 \otimes \mathcal{B}_2$, $A_{x_2} = \{x_1,\ (x_1, x_2) \in A\} \in \mathcal{B}_1$ et $A_{x_1} = \{x_2,\ (x_1, x_2) \in A\} \in \mathcal{B}_2$. On en déduit facilement que, si $f \in (\mathcal{B}_1 \otimes \mathcal{B}_2)^+$ et si μ_i est une mesure sur $(E_i, \mathcal{B}_i)$, $x_1 \mapsto \int f(x_1, x_2)\,d\mu_2(x_2)$ est $\mathcal{B}_1$-mesurable et $x_2 \mapsto \int f(x_1, x_2)\,d\mu_1(x_1)$ est $\mathcal{B}_2$-mesurable.

Théorème 3.5.1. *Soient* $(E_1, \mathcal{B}_1, \mu_1)$ *et* $(E_2, \mathcal{B}_2, \mu_2)$ *deux espaces mesurés avec* μ_1 *et* μ_2 σ*-finies. Il existe une unique mesure sur* $\mathcal{B}_1 \otimes \mathcal{B}_2$*, notée* $\mu_1 \otimes \mu_2$ *et appelée mesure produit de* μ_1 *et* μ_2*, telle que,*

$$\text{pour tous } A_1 \in \mathcal{B}_1,\ A_2 \in \mathcal{B}_2,\ \ \mu_1 \otimes \mu_2(A_1 \times A_2) = \mu_1(A_1)\,\mu(A_2).$$

De plus, pour toute $f \in (\mathcal{B}_1 \otimes \mathcal{B}_2)^+$*,*

$$\int f\,d\mu_1 \otimes \mu_2 = \int [\int f(x_1, x_2)\,d\mu_1(x_1)]\,d\mu_2(x_2) = \int [\int f(x_1, x_2)\,d\mu_2(x_2)]\,d\mu_1(x_1).$$

Preuve: (i) Unicité. On applique la prop. 3.2.2 à $\mathcal{C} = \{A,\ A = A_1 \times A_2,\ A_1 \in \mathcal{B}_1,\ A_2 \in \mathcal{B}_2,\ \mu(A_1) < +\infty,\ \mu(A_2) < +\infty\}$.

(ii) Existence. On applique la prop. 3.4.1 à $I_1(f) = \int [\int f(x_1, x_2)\,d\mu_1(x_1)]\,d\mu_2(x_2)$ ce qui donne l'existence. Mais on peut aussi appliquer la prop. 3.4.1 à $I_2(f) = \int [\int f(x_1, x_2)\,d\mu_2(x_2)]\,d\mu_1(x_1)$ et, vu l'unicité, on a $I_1(f) = I_2(f)$. $\diamond$

Si $f \in \mathcal{L}^1_{\mathbb{C}}(\mu_1 \otimes \mu_2)$, on peut appliquer le théorème précédent à $[\Re(f)]^+$, $[\Re(f)]^-$, $[\Im(f)]^+$ et $[\Im(f)]^-$ et l'on obtient le théorème de Fubini:

Théorème 3.5.2. *Soit* $f \in \mathcal{L}^1_{\mathbb{C}}(\mu_1 \otimes \mu_2)$*. Alors,* $\int |f(x_1, x_2)|\,d\mu_2(x_2) < +\infty$ μ_1 *p.p.,* $\int |f(x_1, x_2)|\,d\mu_1(x_1) < +\infty$ μ_2 *p.p. et, posant* $\phi_1(x_1) = \int f(x_1, x_2)\,d\mu_2(x_2)$*,* $\phi_2(x_2) = \int f(x_1, x_2)\,d\mu_1(x_1)$*,* $\phi_1 \in \mathcal{L}^1(\mu_1)$*,* $\phi_2 \in \mathcal{L}^1(\mu_2)$ *et*

$$\int f\,d\mu_1 \otimes \mu_2 = \int \phi_2(x_2)\,d\mu_2(x_2) = \int \phi_1(x_1)\,d\mu_1(x_1).$$

3.5.2. Tout ceci s'étend sans (trop de) peine au cas de n espaces mesurables. Il y a quelques vérifications fastidieuses à faire du type $\mu_1 \otimes (\mu_2 \otimes \mu_3) = (\mu_1 \otimes \mu_2) \otimes \mu_3$. De plus dans la formule d'intégrations successives, les variables peuvent être intégrées dans tous les ordres possibles. A ce sujet, le grand principe est: soit f mesurable, si f est positive, tout est permis, si f est de signe quelconque ou complexe, on considère d'abord $|f|$ et on commence par montrer que $|f|$ est intégrable.

3.5.3. Mesures de Lebesgue sur $\mathbb{R}^d$.

Lemme 3.5.3. $\mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}) \otimes \ldots \otimes \mathcal{B}(\mathbb{R}) = \mathcal{B}(\mathbb{R}^d)$

Preuve: Soit $\mathcal{B}^{\otimes d} = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}) \otimes \ldots \otimes \mathcal{B}(\mathbb{R})$.

(i) Si est U un ouvert de $\mathbb{R}^d$, $U = \cup_n P_n$, P_n pavé ouvert (i.e. $P_n = \prod_{k=1}^d]a_k, b_k[$). Donc $U \in \mathcal{B}^{\otimes d}$ et $\mathcal{B}(\mathbb{R}^d) \subset \mathcal{B}^{\otimes d}$.

(ii) Soient $X_1, X_2, \ldots, X_d$ les projections canoniques de $\mathbb{R}^d$ sur $\mathbb{R}$. Les X_k sont continues donc mesurable de $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ d'où $\mathcal{B}^{\otimes d} = \sigma(X_1, \ldots, X_d) \subset \mathcal{B}(\mathbb{R}^d)$. $\diamond$

Soit λ la mesure de Lebesgue sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. On définit alors, sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, $\lambda_d = \lambda \otimes \lambda \otimes \ldots \otimes \lambda$. On peut appliquer la prop. 3.2.2 à

$$\mathcal{C} = \{A,\ A = \prod_{i=1}^d]a_i, b_i[,\ -\infty < a_i < b_i < +\infty\}.$$

On obtient que λ_d est l'unique mesure sur $\mathcal{B}(\mathbb{R}^d)$ telle que, pour tous $-\infty < a_i < b_i < +\infty$,

$$\lambda_d(\prod_{i=1}^{d}]a_i, b_i[) = \prod_{i=1}^{d} (b_i - a_i).$$

On appelle λ_d la mesure de Lebesgue sur $\mathbb{R}^d$.

3.5.4. Produit de convolution. Soient μ, ν deux mesures bornées sur $\mathbb{R}^d$. On pose, pour $f \in \mathcal{B}^+(\mathbb{R}^d)$, $I(f) = \int f(x+y)\, d\mu \otimes \nu(x,y)$. On vérifie facilement que $f \mapsto I(f)$ satisfait les hypothèses de la prop. 3.4.1. Il existe donc une unique mesure sur $\mathcal{B}(\mathbb{R}^d)$, notée $\mu * \nu$ et appelée produit de convolution de μ et ν, telle que

$$\int f(x)\, d(\mu * \nu)(x) = \int \int f(x+y)\, d\mu(x) d\nu(y), \quad f \in \mathcal{B}^+(\mathbb{R}^d). \tag{3.9}$$

Propriétés.

(i) $(\mu * \nu)(\mathbb{R}^d) = \mu(\mathbb{R}^d)\nu(\mathbb{R}^d)$,

(ii) $\mu * \nu = \nu * \mu$, $(\mu * \nu) * \rho = \mu * (\nu * \rho)$,

(iii) Si $\mu = \phi.\lambda$, $\nu = \psi.\lambda$ (λ mesure de Lebesgue sur $\mathbb{R}^d$), on a $\mu * \nu = (\phi * \psi).\lambda$ avec

$$\phi * \psi(x) = \int \phi(x-y)\psi(y)\, dy. \tag{3.10}$$

3.5.5. On termine ce chapitre par un résultat très utile. On note C_k l'espace des applications continues à support compact de $\mathbb{R}^d$ dans $\mathbb{R}$ et C_0 l'espace des applications continues de $\mathbb{R}^d$ dans $\mathbb{R}$ tendant vers 0 à l'infini. On munit C_0 de la norme de la convergence uniforme $||f|| = \sup_x |f(x)|$. Rappelons qu'une partie H de C_0 est totale dans C_0 si l'espace vectoriel engendré par H est dense dans $(C_0, ||\ ||)$.

Proposition 3.5.4. *Soient μ, ν deux mesures bornées sur $\mathcal{B}(\mathbb{R}^d)$. On a $\mu = \nu$ dès que l'une des conditions suivantes est satisfaite:*
(i) $\forall\, a_i, b_i \in \mathbb{R}$, $a_i < b_i$, $\mu(]a_1, b_1[\times \ldots \times]a_d, b_d[) = \nu(]a_1, b_1[\times \ldots \times]a_d, b_d[)$,
(ii) $\forall\, f_i \in C_k^+$, $\int f_1(x_1) \ldots f_d(x_d)\, d\mu(x_1, \ldots, x_d) = \int f_1(x_1) \ldots f_d(x_d)\, d\nu(x_1, \ldots, x_d)$.
(iii) il existe un ensemble H total dans C_0 tel que,$\forall\, f \in H$, $\int f\, d\mu = \int f\, d\nu$.

Preuve: Supposons (i) et soit $\mathcal{C} = \{A \in \mathcal{B}(\mathbb{R}^d),\ A =]a_1, b_1[\times \ldots \times]a_d, b_d[\}$. $\mathcal{C}$ est stable par intersection finie et $\sigma(\mathcal{C}) = \mathcal{B}(\mathbb{R}^d)$. Donc (cor. 3.2.3) $\mu = \nu$.

Supposons (ii). Puisque, pour tous $a < b$, $1_{]a,b[} = \lim \uparrow f_n$ avec $f_n \in C_k^+$, (ii) implique (i) (convergence monotone) et le résultat cherché.

Supposons (iii) et soit $V = $ e.v.$[H]$. On a, pour toute $f \in V$, $\int f\, d\mu = \int f\, d\nu$. Soient $f \in C_0$ et $f_n \in V$ tendant vers f dans $(C_0, ||\ ||)$. Vu que $|\int f_n\, d\mu - \int f\, d\mu| \leq ||f_n - f||\, \mu(\mathbb{R}^d)$, $\int f_n\, d\mu \to_n \int f\, d\mu$. De même $\int g_n\, d\nu \to_n \int g\, d\nu$ d'où $\int f\, d\mu = \int f\, d\nu$ pour toute $f \in C_0$. On applique (ii). $\diamond$

Pour montrer qu'une partie de C_0 est dense, le théorème de Stone-Weierstrass est un outil précieux. Rappelons qu'une sous-algèbre V de C_0 est un sous-espace vectoriel tel que $f, g \in V$ implique $fg \in V$. Alors:

Théorème 3.5.5. *Soit A une sous-algèbre de C_0 vérifiant*
(i) pour tous $x, y \in \mathbb{R}^d$, $x \neq y$, il existe $f \in A$ telle que $f(x) \neq f(y)$,
(ii) pour tout $x \in \mathbb{R}^d$, il existe $f \in A$ telle que $f(x) \neq 0$,
alors $\overline{A} = C_0$.

Notant C_k^∞ l'espace des fonctions indéfiniment dérivables à support compact sur $\mathbb{R}^d$, on a:

Corollaire 3.5.6. *C_k^∞ est dense dans C_0.*

Preuve: Soit, pour $t \in R$, $\phi(t) = 1_{]0,+\infty[}(t) \exp(-\frac{1}{t^2})$. On vérifie facilement que $\phi \in C^\infty(\mathbb{R})$. On pose, pour $\rho > 0$, $a \in \mathbb{R}^d$ et $x \in \mathbb{R}^d$, $f_{\rho,a}(x) = \phi(\rho^2 - |x-a|^2)$. On a $f_{\rho,a} \in C_k^\infty$, $f_{\rho,a}(a) > 0$, $f_{\rho,a}(x) = 0$ si $|x-a| > \rho$. On peut alors appliquer le th. 3.5.5

Chapitre 4

Espace de probabilité général. Variables aléatoires

4.1. Espace de probabilité

4.1.1. On peut maintenant aborder le cas général.

Définition 4.1.1. *On appelle espace de probabilité un triplet* $(\Omega, \mathcal{A}, \mathbb{P})$ *où* $(\Omega, \mathcal{A})$ *est un espace mesurable et* $\mathbb{P}$ *une probabilité sur* $\mathcal{A}$.

Les éléments de $\mathcal{A}$ s'appellent des événements. Pour des événements A et B, on écrira indifféremment $A \cap B$ ou AB.

Premières propriétés. A_n, A, B étant des événements,

(i) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$; si $A \subset B$, $\mathbb{P}(A) \leq \mathbb{P}(B)$,

(ii) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$,

(iii) si $A_n \uparrow A$, $\mathbb{P}(A_n) \uparrow \mathbb{P}(A)$,

(iv) si $A_n \downarrow A$, $\mathbb{P}(A_n) \downarrow \mathbb{P}(A)$,

(v) $\mathbb{P}(\cup A_n) \leq \sum \mathbb{P}(A_n)$.

Rappelons qu'un sous-ensemble B de Ω est dit négligeable si $B \subset A \in \mathcal{A}$ tel que $\mathbb{P}(A) = 0$. Une propriété dépendant de ω est vraie presque sûrement, en abrégé p.s., si elle est vraie en dehors d'un ensemble négligeable. Notons qu'un ensemble négligeable n'est pas toujours un événement sauf si l'espace $(\Omega, \mathcal{A}, \mathbb{P})$ est complet. On peut cependant toujours se ramener à ce cas. Voir à ce sujet 3.2.2.

4.1.2. Probabilité conditionnelle. Toutes les définitions et résultats de la section 1.3 restent valables en supposant que tous les ensembles considérés sont des événements i.e. sont des éléments de $\mathcal{A}$. En particulier la définition de n événements indépendants (def. 1.3.5) est inchangée. On dit alors que des événements $(A_n)_{n \in \mathbb{N}}$ sont indépendants si, pour tout r, $A_1, \ldots, A_r$ sont indépendants.

4.1.3. Lemme de Borel-Cantelli. On appelle traditionnellement ainsi le point (i) de la proposition suivante; (ii) s'appelant la réciproque du lemme de Borel-Cantelli.

Etant donné une suite $(A_n,\, n \in \mathbb{N})$ d'événements, on pose:

$$\limsup A_n = \cap_n \cup_{k\geq n} A_k = \lim \downarrow_n \cup_{k\geq n} A_k.$$

On a donc $\limsup A_n = \{\omega,\ \omega \in A_n$ pour une infinité de $n\} = \{\sum_n 1_{A_n} = +\infty\}$ et $1_{\limsup A_n} = \limsup 1_{A_n}$, ce qui justifie la dénomination.

Proposition 4.1.2. *Soit $(A_n,\, n \geq 0)$ une suite d'événements.*
(i) Si $\sum_n \mathbb{P}(A_n) < +\infty$, $\mathbb{P}(\limsup A_n) = 0$.
(ii) Si les A_n sont indépendants et si $\sum_n \mathbb{P}(A_n) = +\infty$, $\mathbb{P}(\limsup A_n) = 1$.

Preuve: (i) On a

$$\mathbb{P}(\limsup A_n) = \lim \downarrow_n \mathbb{P}(\cup_{k\geq n} A_k) \leq \lim \downarrow_n \sum_{k=n}^{\infty} \mathbb{P}(A_k) = 0.$$

(ii) Vu l'inégalité $1 - u \leq \mathrm{e}^{-u}$ et l'indépendance des A_n^c, on a

$$\mathbb{P}(\cap_{k=n}^m A_k^c) = \prod_{k=n}^m \mathbb{P}(A_k^c) = \prod_{k=n}^m (1 - \mathbb{P}(A_k)) \leq \exp(-\sum_{k=n}^m \mathbb{P}(A_k))$$
$$\text{donc } \mathbb{P}(\cap_{k=n}^\infty A_k^c) = \lim \downarrow_m \mathbb{P}(\cap_{k=n}^m A_k^c) = 0 \text{ si } \sum \mathbb{P}(A_n) = +\infty.$$

Passant au complémentaire, on a $\mathbb{P}(\cup_{k=n}^\infty A_k) = 1$ et $\mathbb{P}(\limsup A_n) = 1$. $\diamond$

4.2. Variables aléatoires

4.2.1. Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité et $(E, \mathcal{E})$ un espace mesurable.

Définition 4.2.1. *On appelle variable aléatoire (en abrégé v.a.) à valeurs $(E, \mathcal{E})$ toute application mesurable de $(\Omega, \mathcal{A})$ dans $(E, \mathcal{E})$.*

Si E est dénombrable et $\mathcal{E} = \mathcal{P}(E)$, on parle de v.a. discrète,

si $E = \overline{\mathbb{R}^+}$ et $\mathcal{E} = \mathcal{B}(\overline{\mathbb{R}^+})$, on parle de v.a. positive,

si $E = \mathbb{R}$ et $\mathcal{E} = \mathcal{B}(\mathbb{R})$, on parle de v.a. réelle (v.a.r.),

si $E = \mathbb{R}^d$ et $\mathcal{E} = \mathcal{B}(\mathbb{R}^d)$, on parle de v.a. vectorielle,

si $E = \mathbb{C}$ et $\mathcal{E} = \mathcal{B}(\mathbb{C})$, on parle de v.a. complexe.

4.2.2. Loi d'une v.a.. Soient X une v.a. à valeurs $(E, \mathcal{E})$ et $\Gamma \in \mathcal{E}$. Rappelons qu'on note

$$\{X \in \Gamma\} = \{\omega,\ X(\omega) \in \Gamma\} = X^{-1}(\Gamma).$$

On pose alors:

$$\mu_X(\Gamma) = \mathbb{P}(X \in \Gamma), \;\; \Gamma \in \mathcal{E}. \tag{4.1}$$

Evidemment $\mu_X(\Gamma) \leq 1$ et $\mu_X(E) = 1$. Soient $\Gamma_n \in \mathcal{E}$ des ensembles deux à deux disjoints. Vu que

$$X^{-1}(\Gamma_m \cap \Gamma_n) = X^{-1}(\Gamma_m) \cap X^{-1}(\Gamma_n), \quad X^{-1}(\cup_n \Gamma_n) = \cup_n X^{-1}(\Gamma_n),$$

les ensembles $X^{-1}(\Gamma_n)$ sont deux à deux disjoints d'union $X^{-1}(\cup_n \Gamma_n)$. On a donc

$$\mu_X(\cup_n \Gamma_n) = \mathbb{P}(X^{-1}(\cup_n \Gamma_n)) = \sum_n \mathbb{P}(X^{-1}(\Gamma_n)) = \sum_n \mu_X(\Gamma_n).$$

Ceci montre que μ_X est une probabilité sur $(E, \mathcal{E})$.

Définition 4.2.2. *Soit X une v.a. à valeurs $(E, \mathcal{E})$. La probabilité μ_X définie par (4.1) s'appelle la loi de X.*

4.2.3. Espérance.

Définition 4.2.3. *(i) Soit X une v.a. positive. On appelle espérance de X et on note $\mathbb{E}(X)$ la quantité $\int X d\mathbb{P}$.*
(ii) Soit X une v.a. complexe telle que $\mathbb{E}(|X|) < +\infty$. On appelle espérance de X et on note $\mathbb{E}(X)$ la quantité $\int X d\mathbb{P}$.

Vu (3.3), on a pour toute v.a. positive X,

$$\mathbb{E}(X) = \lim \uparrow \sum_{k=0}^{n2^n-1} \frac{k}{2^n} \mathbb{P}(\frac{k}{2^n} \leq X < \frac{k+1}{2^n}) + n\mathbb{P}(X \geq n). \tag{4.2}$$

Plus généralement, soient X une v.a. à valeurs $(E, \mathcal{E})$ et $f : E \to \mathbb{R}$ $\mathcal{E}$-mesurable, alors $f(X)$ est une v.a. réelle et on peut considérer $\mathbb{E}(f(X))$ si $f \geq 0$ ou si $\mathbb{E}(|f(X)|) < +\infty$. Alors,

Théorème 4.2.4. *Soit X une v.a. à valeurs $(E, \mathcal{E})$ de loi μ_X, on a,*

$$\textit{pour toute } f \in \mathcal{E}^+ \cup \mathcal{L}^1(E, \mathcal{E}, \mu_X), \; \mathbb{E}(f(X)) = \int f \, d\mu_X \, . \tag{4.3}$$

Preuve: Si $f = 1_\Gamma$, c'est la définition de μ_X. Donc (4.3) est vraie pour f étagée puis (limite croissante) pour $f \in \mathcal{E}^+$. Enfin, pour $f \in \mathcal{L}^1(E, \mathcal{E}, \mu_X)$, il suffit d'écrire $f = f^+ - f^-$.

Exemples. Il y a deux situations fondamentales.

(i) X est discrète i.e. E est dénombrable. La loi μ_X est alors déterminée par la famille $(\mu_X(a), \; a \in E)$ où $\mu_X(a) := \mu_X(\{a\}) = \mathbb{P}(X = a)$ et l'on a

$$\text{pour toute } f \geq 0, \quad \mathbb{E}(f(X)) = \sum_{a \in E} f(a)\mu_X(a). \tag{4.4}$$

(ii) X est vectorielle i.e. à valeurs $\mathbb{R}^d$ et $\mu_X = h_X.\lambda$, λ étant la mesure de Lebesgue sur $\mathbb{R}^d$ (3.5.3). On dit alors que X est une v.a. de densité h_X. Dans ce cas, on a,

$$\text{pour toute } f \in \mathcal{B}^+(\mathbb{R}^d), \quad \mathbb{E}(f(X)) = \int f h_X \, d\lambda. \tag{4.5}$$

4.2.4. Moments. Dans la suite L^p désigne $L^p(\Omega, \mathcal{A}, \mathbb{P})$. On ne distinguera pas deux v.a.r. égales p.s. ce qui fait qu'on désigne par X aussi bien la v.a. X que sa classe d'équivalence dans L^0. En particulier on écrira indifféremment $X \in L^p$ aussi bien que $X \in \mathcal{L}^p$. Notons que, si $1 \le q \le p$, $L^p \subset L^q$ puisque $|X|^q \le 1 + |X|^p$. En fait, d'après (3.7), on a le résultat plus précis:

$$\{\mathbb{E}(|X|^q)\}^{1/q} \le \{\mathbb{E}(|X|^p)\}^{1/p}, \quad q \le p.$$

Définition 4.2.5. *Soit X une v.a.r. Pour $p \in [1, +\infty[$, $\mathbb{E}|X|^p$ s'appelle moment absolu d'ordre p de X; pour $p \in \mathbb{N}^*$, si $X \in L^p$, $\mathbb{E}(X^p)$ s'appelle moment d'ordre p de X.*

Notons que, d'après (4.3), $\mathbb{E}(|X|^p) = \int |x|^p \, d\mu_X(x)$, $\mathbb{E}(X^p) = \int x^p \, d\mu_X(x)$. Les deux moments les plus importants sont le moment d'ordre 1 qui n'est rien d'autre que l'espérance de X (on dit aussi la moyenne de X) et le moment d'ordre 2. On pose, pour $X \in L^2$,

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2] \tag{4.6}$$

qu'on appelle la variance de X. On a $\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$ et:

Lemme 4.2.6. *Si $Y \in L^2$, $\mathbb{E}[(Y-a)^2]$ est minimum pour $a = \mathbb{E}(Y)$ et ce minimum vaut Var(Y).*

Preuve: En effet, si $m = \mathbb{E}(Y)$, $\mathbb{E}[(Y-a)^2] = \mathbb{E}[(Y-m)^2] + (m-a)^2$. $\diamond$

On note aussi σ_X^2 pour $\mathrm{Var}(X)$, la racine carrée positive de $\mathrm{Var}(X)$ s'appelle l'écart type et se note σ_X. Une v.a. $X \in L^1$ est dite centrée si $\mathbb{E}(X) = 0$. Une v.a. $X \in L^2$ est dite centrée réduite si $\mathbb{E}(X) = 0$ et $\mathbb{E}(X^2) = \mathrm{Var}(X) = 1$. Noter que, si $X \in L^2$ et $\sigma_X > 0$, $\sigma_X^{-1}(X - \mathbb{E}(X))$ est centrée réduite.

Proposition 4.2.7. *(i) Soit $X \in L^p$, $p \ge 1$. On a, pour tout $\lambda > 0$,*

$$\mathbb{P}(|X| \ge \lambda) \le \frac{1}{\lambda^p} \mathbb{E}|X|^p.$$

(ii) Soit $X \in L^2$. On a, pour tout $\lambda > 0$,

$$\mathbb{P}(|X - \mathbb{E}(X)| \ge \lambda) \le \frac{1}{\lambda^2} \mathit{Var}(X).$$

Preuve: (i) On remarque que $\lambda^p 1_{\{|X| \ge \lambda\}} \le |X|^p$ et on prend l'espérance.

(ii) On applique (i) à $|X - \mathbb{E}(X)|$. $\diamond$

La première de ces inégalités s'appellent l'inégalité de Markov, la seconde l'inégalité de Bienaymé-Tchebichev. Montrons maintenant l'inégalité de Jensen.

Proposition 4.2.8. *Soient X une v.a.r. et f une application convexe de $\mathbb{R}$ dans $\mathbb{R}$. On suppose X et $f(X)$ intégrables. Alors $f(\mathbb{E}(X)) \leq \mathbb{E}(f(X))$.*

Preuve: Soit $m = \mathbb{E}(X)$. La fonction f étant convexe, il existe une droite passant par $(m, f(m))$ et située sous le graphe de f i.e. une fonction affine $\alpha(x) = a(x-m) + f(m) \leq f(x)$ pour tout $x \in \mathbb{R}$. On a donc $a(X-m) + f(m) \leq f(X)$ et, prenant l'espérance, $f(m) \leq \mathbb{E}(f(X))$. ◇

Corollaire 4.2.9. *Soient μ une probabilité sur $\mathbb{R}$, f une application convexe de $\mathbb{R}$ dans $\mathbb{R}$ et $g \in [\mathcal{B}(\mathbb{R})]$. On suppose g et $f \circ g$ μ-intégrables. Alors*

$$f(\int g(x)\, d\mu(x)) \leq \int f(g(x))\, d\mu(x).$$

Preuve: On choisit $\Omega = \mathbb{R}$, $\mathcal{A} = \mathcal{B}(\mathbb{R})$, $\mathbb{P} = \mu$, $X = g$ et on applique la prop. 4.2.8. ◇

4.3. Probabilités sur $\mathbb{R}$

4.3.1. On a vu en 2.2 des exemples de lois discrètes sur $\mathbb{R}$. On considère maintenant quelques lois à densités. Une application borélienne q de $\mathbb{R}^d$ dans $\mathbb{R}$ est une densité de probabilité si:

$$q(x) \geq 0, \quad \int_{\mathbb{R}^d} q(x)\, dx = 1. \tag{4.7}$$

On dit alors qu'une v.a. à valeurs $\mathbb{R}^d$ X a pour densité $q(x)$ si la loi de X est de densité q par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$ ce qu'on écrit $\mu_X = q.\lambda$. Dans cette section, on suppose $d = 1$.

a. Loi uniforme sur $[a, b]$ notée $U(a, b)$, $a, b \in \mathbb{R}$. C'est la loi sur $\mathbb{R}$ de densité

$$q(x) = \frac{1}{b-a} 1_{[a,b]}(x). \tag{4.8}$$

Si $X \sim U(a, b)$, $\mathbb{E}(X) = \frac{a+b}{2}$, $\mathrm{Var}(X) = \frac{(b-a)^2}{12}$.

b. Loi de Cauchy de paramètre $a > 0$. C'est la loi de densité

$$q_a(x) = \frac{1}{\pi(1 + (x-a)^2)}. \tag{4.9}$$

Noter que, si X suit une loi de Cauchy, $\mathbb{E}(|X|) = +\infty$.

c. Loi de Laplace. C'est la loi de densité

$$q(x) = \frac{1}{2} \mathrm{e}^{-|x|}. \tag{4.10}$$

Noter que, si X suit une loi de Laplace, $\mathbb{E}(X) = 0$, $\mathbb{E}(X^2) = 2$.

d. Loi gamma de paramètres a, c, $a > 0$, $c > 0$, notée $G(a, c)$. Rappelons que la fonction

$$\Gamma(a) = \int_0^{+\infty} \mathrm{e}^{-x} x^{a-1}\, dx \tag{4.11}$$

est définie pour tout $a > 0$ et que l'on a $\Gamma(1) = 1$, $\Gamma(a+1) = a\Gamma(a)$ (intégrer par parties) d'où $\Gamma(n) = (n-1)!$. Donc

$$q_{a,c}(x) = \frac{c^a}{\Gamma(a)} \mathrm{e}^{-cx} x^{a-1} 1_{\mathbb{R}^+}(x) \tag{4.12}$$

est une densité de probabilité sur $\mathbb{R}$. La loi de densité $q_{a,c}$ s'appelle la loi $G(a, c)$. On a, si $X \sim G(a, c)$, $\mathbb{E}(X) = a/c$, $\mathrm{Var}(X) = a/c^2$.

En particulier, pour $a = 1$, on obtient la loi $G(1, c)$ de densité $c\mathrm{e}^{-cx}$ qu'on appelle loi exponentielle de paramètre c.

e. Loi normale ou de Gauss $N_1(m, \sigma^2)$. On appelle loi $N_1(m, \sigma^2)$ la loi sur $\mathbb{R}$ de densité

$$f_{m,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(x-m)^2}{2\sigma^2}}. \tag{4.13}$$

Si $X \sim N_1(m, \sigma^2)$, $\mathbb{E}(X) = m$, $\mathrm{Var}(X) = \sigma^2$. Noter que si $X \sim N_1(0, 1)$, $m + \sigma X \sim N_1(m, \sigma^2)$.

4.3.2. Fonction de répartition. On a vu en 3.2.3 que, si μ est une probabilité sur $\mathbb{R}$, la fonction $F(t) = \mu(]-\infty, t])$ est croissante de 0 à 1 et continue à droite et que, réciproquement, si une fonction F a ces propriétés, il existe une probabilité μ sur $\mathbb{R}$, unique, telle que $F(t) = \mu(]-\infty, t])$. La fonction F s'appelle la fonction de répartition de μ.

Définition 4.3.1. *Soit X une v.a. réelle de loi μ_X. On appelle fonction de répartition de X la fonction*

$$F_X(t) = \mathbb{P}(X \le t) = \mu_X(]-\infty, t]).$$

Il résulte du rappel que F_X croît de 0 à 1 et est continue à droite. Elle a donc une limite à gauche en tout point notée $F_X(x-)$. De plus, on a

$$\mathbb{P}(a < X \le b) = \mathbb{P}(X \le b) - \mathbb{P}(X \le a) = F_X(b) - F_X(a).$$

En particulier $\mathbb{P}(a - \varepsilon < X \le a) = F_X(a) - F_X(a - \varepsilon)$ d'où, lorsque $\varepsilon \downarrow 0$,

$$\mu_X(\{a\}) = \mathbb{P}(X = a) = F_X(a) - F_X(a-).$$

Etant donnée une fonction de répartition F, on pose, pour $u \in [0, 1]$,

$$F^{-1}(u) = \inf(t,\ F(t) \ge u). \tag{4.14}$$

Proposition 4.3.2. *Soit μ une probabilité sur $\mathbb{R}$ de fonction de répartition F et U une v.a.r. de loi uniforme sur $[0, 1]$. Alors $F^{-1}(U)$ est une v.a. de loi μ.*

Preuve: Considérons, pour $u \in [0,1]$ fixé, $I(u) = \{t, F(t) \geq u)$. Puisque F est croissante, c'est un intervalle de la forme $[F^{-1}(u), +\infty[$ ou $]F^{-1}(u), +\infty[$. Soit $t_n \downarrow F^{-1}(u)$. Alors $F(t_n) \geq u$ et (continuité à droite) $F(F^{-1}(u)) \geq u$ i.e. $F^{-1}(u) \in I(u) = [F^{-1}(u), +\infty[$. On a donc

$$\{u \leq F(t)\} \Leftrightarrow \{t \geq F^{-1}(u)\}. \tag{4.15}$$

Finalement

$$\mathbb{P}(F^{-1}(U) \leq t) = \mathbb{P}(U \leq F(t)) = F(t).$$

En conclusion, $X = F^{-1}(U)$ a pour fonction de répartition F i.e. a pour loi μ. $\diamond$

4.4. Variables aléatoires indépendantes

4.4.1. Dans cette sous-section, $X_1, \ldots, X_n$ désignent des v.a. à valeurs $(E_1, \mathcal{E}_1), \ldots, (E_n, \mathcal{E}_n)$.

Définition 4.4.1. *Les v.a. $X_1, \ldots, X_n$ sont dites indépendantes si:*

$$\textit{pour tous } \Gamma_k \in \mathcal{E}_k,\ \mathbb{P}(X_1 \in \Gamma_1, \ldots, X_n \in \Gamma_n) = \mathbb{P}(X_1 \in \Gamma_1) \ldots \mathbb{P}(X_n \in \Gamma_n). \tag{4.16}$$

La suite $(X_n,\ n \in \mathbb{N})$ est dite indépendante si, pour tout n, les v.a. $X_1, \ldots, X_n$ sont indépendantes.

Supposons $n = 2$. On peut considérer (X_1, X_2) comme une v.a. à valeurs $(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$. Sa loi est alors définie par

$$\mu_{(X_1,X_2)}(\Gamma_1 \times \Gamma_2) = \mathbb{P}(X_1 \in \Gamma_1, X_2 \in \Gamma_2).$$

Il résulte donc du th. 3.5.1 que X_1 et X_2 sont indépendantes ssi $\mu_{(X_1,X_2)} = \mu_{X_1} \otimes \mu_{X_2}$. Il en est de même pour n quelconque et on peut énoncer:

Proposition 4.4.2. *Les v.a. $X_1, \ldots, X_n$ sont indépendantes ssi $\mu_{(X_1,\ldots,X_n)} = \mu_{X_1} \otimes \ldots \otimes \mu_{X_n}$.*

Le résultat suivant, un peu technique, est très utile.

Proposition 4.4.3. *Soit $\mathcal{C}_k \subset \mathcal{E}_k$ une classe contenant E_k, stable par intersection finie, et telle que $\sigma(\mathcal{C}_k) = \mathcal{E}_k$, $k = 1, \ldots, n$. Si*

$$\textit{pour tous } \Gamma_k \in \mathcal{C}_k,\ \mathbb{P}(X_1 \in \Gamma_1, \ldots, X_n \in \Gamma_n) = \mathbb{P}(X_1 \in \Gamma_1) \ldots \mathbb{P}(X_n \in \Gamma_n),$$

les v.a. $X_1, \ldots, X_n$ sont indépendantes.

Preuve: Soit $\mathcal{C} = \{\Gamma,\ \Gamma = \Gamma_1 \times \ldots \times \Gamma_n,\ \Gamma_k \in \mathcal{C}_k\}$. Alors $\mathcal{C}$ est stable par intersection finie et $\sigma(\mathcal{C}) = \mathcal{E}_1 \otimes \ldots \otimes \mathcal{E}_n$ (en effet $E_1 \times \ldots \times E_{k-1} \times \Gamma_k \times E_{k+1} \times \ldots \times E_n \in \mathcal{C}$ si $\Gamma_k \in \mathcal{C}_k$ et donc $E_1 \times \ldots \times E_{k-1} \times \Gamma_k \times E_{k+1} \times \ldots \times E_n \in \sigma(\mathcal{C})$ si $\Gamma_k \in \mathcal{E}_k$). Par hypothèse, pour tout $\Gamma \in \mathcal{C}$, $\mu_{(X_1,\ldots,X_n)}(\Gamma) = \mu_{X_1} \otimes \ldots \otimes \mu_{X_n}(\Gamma)$. Donc (prop. 3.2.2) $\mu_{(X_1,\ldots,X_n)} = \mu_{X_1} \otimes \ldots \otimes \mu_{X_N}$ et les v.a.$X_1, \ldots, X_n$ sont indépendantes. $\diamond$

Théorème 4.4.4. *Les v.a.* $X_1, \ldots, X_n$ *sont indépendantes ssi, pour toutes* $f_i \in \mathcal{E}_i^+$,

$$\mathbb{E}(f_1(X_1) \ldots f_n(X_n)) = \mathbb{E}(f_1(X_1)) \ldots \mathbb{E}(f_n(X_n)). \tag{4.17}$$

Dans ce cas, si, pour $k = 1, 2, \ldots, n$, $\mathbb{E}(|f_k(X_k)|) < +\infty$, *on a* $\mathbb{E}(|f_1(X_1) \ldots f_n(X_n)|) < +\infty$ *et (4.17) est satisfaite.*

Preuve: On suppose $n = 2$.

(i) Si on a (4.17), il suffit de choisir $f_1 = 1_{\Gamma_1}, f_2 = 1_{\Gamma_2}$ pour avoir l'indépendance de X_1 et X_2.

(ii) Supposons X_1 et X_2 indépendantes. On a, pour $f_k \in \mathcal{E}_k^+$, $k = 1, 2$,

$$\mathbb{E}(f_1(X_1)f_2(X_2)) = \int f_1(x_1)f_2(x_2)\, d\mu_{(X_1,X_2)}(x_1, x_2) = \int f_1(x_1)f_2(x_2)\, d\mu_{X_1}(x_1)d\mu_{X_2}(x_2)$$
$$= \int f_1(x_1)\, d\mu_{X_1}(x_1) \int f_2(x_2)\, d\mu_{X_2}(x_2) = \mathbb{E}(f_1(X_1))\mathbb{E}(f_2(X_2)).$$

Enfin si $\mathbb{E}(|f_k(X_k)|) < +\infty$, $k = 1, 2$,

$$\mathbb{E}(|f_1(X_1)f_2(X_2)|) = \mathbb{E}(|f_1(X_1)|)\mathbb{E}(|f_2(X_2)|) < +\infty$$

et le calcul ci-dessus reste valable. $\diamond$

On en déduit facilement, comme en 2.2.6, que, si les v.a. $X_1, X_2, \ldots, X_n$ sont indépendantes:

a. Pour toute permutation $\{r_1, \ldots, r_n\}$ de $\{1, \ldots, n\}$, les v.a. $X_{r(1)}, \ldots, X_{r(n)}$ sont indépendantes.

b. Pour toutes $g_k \in [\mathcal{E}_k]$, les v.a. $g_1(X_1), \ldots, g_n(X_n)$ sont indépendantes.

c. Posant

$$Y_1 = (X_1, \ldots, X_{r_1}), Y_2 = (X_{r_1+1}, \ldots, X_{r_2}), \ldots, Y_p = (X_{r_{p-1}+1}, \ldots, X_{r_p}),$$

les v.a. $Y_1, \ldots, Y_p$ sont indépendantes.

4.4.2. On s'intéresse plus particulièrement aux v.a. réelles. Les prop. 4.4.2 et 3.5.4 impliquent immédiatement:

Proposition 4.4.5. *Soient* $X_1, \ldots, X_n$ *des v.a. réelles. Il y a équivalence entre:*
(i) Les v.a. $X_1, \ldots, X_n$ *sont indépendantes,*
(ii) $\forall\, a_i < b_i$, $\mathbb{P}(a_i < X_i < b_i,\, i = 1, \ldots, n) = \prod_{i=1}^n \mathbb{P}(a_i < X_i < b_i)$,
(iii) $\forall\, f_i \in C_k^+$, $\mathbb{E}(f_1(X_1) \ldots f_n(X_n)) = \mathbb{E}(f_1(X_1)) \ldots \mathbb{E}(f_n(X_n))$.

4.4.3. Covariance. Soient X et Y deux v.a.r. réelles de carré intégrable. On appelle covariance de X et Y et on note $\mathrm{Cov}(X, Y)$ la quantité

$$\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))] = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y). \tag{4.18}$$

Propriétés.

(i) Cov(X,X)=Var(X). Pour tous $a, b \in \mathbb{R}$, $\mathrm{Cov}(X+a, Y+b) = \mathrm{Cov}(X, Y)$.

(ii) Si les v.a. X et Y sont indépendantes, $\mathrm{Cov}(X, Y) = 0$.

(iii) $(X, Y) \mapsto \mathrm{Cov}(X, Y)$ est une forme bilinéaire symétrique. En particulier, vu (i),

$$\mathrm{Var}(\sum_{k=1}^{n} X_k) = \sum_{k=1}^{n} \mathrm{Var}(X_k) + 2 \sum_{1 \leq j < k \leq n} \mathrm{Cov}(X_j, X_k).$$

Remarque. $\mathrm{Cov}(X, Y) = 0$ n'implique pas l'indépendance de X et Y. Par exemple si la loi du couple (X, Y) est donnée par:

$$\mathbb{P}((X,Y) = (1,0)) = \mathbb{P}((X,Y) = (-1,0)) = \mathbb{P}((X,Y) = (0,1)) = \mathbb{P}((X,Y) = (0,-1)) = \frac{1}{4},$$

on a $\mathbb{E}(X) = \mathbb{E}(Y) = \mathbb{E}(XY) = \mathrm{Cov}(X, Y) = 0$ et $\mathbb{P}(X = 1, Y = 0) = \frac{1}{4} \neq \mathbb{P}(X = 1)\mathbb{P}(Y = 0) = \frac{1}{8}$.

4.4.4. Coefficient de corrélation. Soient X et Y deux v.a. réelles de carré intégrable non p.s. constantes (donc $\mathrm{Var}(X) > 0$, $\mathrm{Var}(Y) > 0$). On appelle coefficient de corrélation de X et Y et on note $\rho(X, Y)$ la quantité

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}. \tag{4.19}$$

Noter que (inégalité de Schwarz) $|\rho(X, Y)| \leq 1$, que $\rho(X, Y) = \rho(Y, X)$ et que $\rho(X, Y) = 0$ si X et Y sont indépendantes. De plus

Proposition 4.4.6. *Soit X et Y deux v.a.r. de carré intégrable non p.s. constantes. Alors $\varepsilon(a, b) = \mathbb{E}(Y - aX - b)^2$ est minimum pour*

$$\hat{a} = \frac{Cov(X,Y)}{Var(X)}, \;\; \hat{b} = \mathbb{E}(Y) - \hat{a}\,\mathbb{E}(X)$$

et ce minimum vaut $Var(Y)(1 - \rho^2(X, Y))$.

Preuve: Posant $\tilde{X} = X - \mathbb{E}(X)$, $\tilde{Y} = Y - \mathbb{E}(Y)$, $\tilde{b} = b - \mathbb{E}(Y) + a\,\mathbb{E}(X)$, on a

$$\begin{aligned} \varepsilon(a,b) &= \mathbb{E}[(\tilde{Y} - a\tilde{X} - \tilde{b})^2] = \mathbb{E}(\tilde{Y}^2) + a^2\,\mathbb{E}(\tilde{X}^2) + \tilde{b}^2 - 2a\,\mathbb{E}(\tilde{X}\tilde{Y}) \\ &= \mathrm{Var}(X)(a - \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)})^2 + \tilde{b}^2 + \mathrm{Var}(Y) - \frac{\mathrm{Cov}^2(X,Y)}{\mathrm{Var}(X)}. \end{aligned}$$

Donc $\varepsilon(a, b)$ est minimum pour $a = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)} = \hat{a}$ et $\tilde{b} = 0$ i.e. $b = \hat{b} = \mathbb{E}(Y) - \hat{a}\,\mathbb{E}(X)$ et ce minimum vaut $\mathrm{Var}(Y) - \frac{\mathrm{Cov}^2(X,Y)}{\mathrm{Var}(X)} = \mathrm{Var}(Y)(1 - \rho^2(X, Y))$. ◇

Cette proposition implique que $|\rho(X, Y)| = 1$ ssi $Y = aX + b$ p.s.

4.5. Vecteurs aléatoires

4.5.1. Notations. (i) On note, pour $x = (x_1, \dots, x_d) \in \mathbb{R}^d$, $|x| = (x_1^2 + \dots + x_d^2)^{1/2}$.

(ii) On note $L_d^p = \{X = (X_1, \dots, X_d),\ X_k$ v.a. réelles et $\mathbb{E}|X|^p < +\infty\}$.

(iii) Si $X \in L_d^1$, on note $\mathbb{E}(X) = (\mathbb{E}(X_1), \dots, \mathbb{E}(X_d))$.

4.5.2. On appelle vecteur aléatoire toute v.a. à valeurs $\mathbb{R}^d$. On remarque d'abord que $X = (X_1, \dots, X_d)$ est un vecteur aléatoire ssi, pour $k = 1, \dots, d$, X_k est une v.a.r. Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire. Les lois $\mu_{X_1}, \dots, \mu_{X_d}$ s'appellent les lois marginales de X.

Proposition 4.5.1. *Soit X un vecteur aléatoire de densité q. Alors X_k a pour densité*

$$q_k(u) = \int q(x_1, \dots, x_{k-1}, u, x_{k+1}, \dots, x_d)\, dx_1 \dots dx_{k-1} dx_{k+1} \dots dx_d.$$

Preuve: On suppose $d = 2$. Alors, pour $\phi \in \mathcal{B}^+(\mathbb{R})$,

$$\mathbb{E}(\phi(X_1)) = \int \phi(x_1) q(x_1, x_2)\, dx_1 dx_2 = \int \phi(x_1) \{\int q(x_1, x_2)\, dx_2\} dx_1. \ \diamond$$

On sait (th. 4.4.2) que les composantes $X_1, \dots, X_d$ sont indépendantes ssi $\mu_X = \mu_{X_1} \otimes \dots \otimes \mu_{X_d}$. On en déduit immédiatement:

Proposition 4.5.2. *Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire de densité q. Les composantes $X_1, \dots, X_d$ sont indépendantes ssi*

$$q(x_1, \dots, x_d) = q_1(x_1) \dots q_d(x_d) \quad p.p.$$

où q_k est la densité de X_k.

En fait pour montrer l'indépendance de $X_1, \dots, X_d$, on utilise plutôt:

Corollaire 4.5.3. *Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire de densité q. Les composantes $X_1, \dots, X_d$ sont indépendantes ssi*

$$q(x_1, \dots, x_d) = g_1(x_1) \dots g_d(x_d) \quad p.p.$$

et alors X_k a pour densité $q_k(u) = g_k(u) / \int_{\mathbb{R}} g_k(v)\, dv$.

Preuve: ($d = 2$) On suppose que $q(x_1, x_2) = g_1(x_1) g_2(x_2)$. La densité q_1 de X_1 est donc

$$q_1(x_1) = \int g_1(x_1) g_2(x_2)\, dx_2 = a_1 g_1(x_1), \quad a_1 = \int g_2(x_2)\, dx_2.$$

De même $q_2(x_2) = a_2 g_2(x_2), \quad a_2 = \int g_1(x_1)\, dx_1$. Mais

$$1 = \int q(x_1, x_2)\, dx_1 dx_2 = \int g_1(x_1) g_2(x_2)\, dx_1 dx_2 = \int g_1(x_1)\, dx_1 \int g_2(x_2)\, dx_2 = a_1 a_2.$$

On conclut facilement. $\diamond$

4.5.3. Matrice de covariance (ou de dispersion). On note M^{T} la matrice transposée de la matrice M. Alors on peut représenter $x \in \mathbb{R}^d$ par un vecteur colonne i.e. une matrice $d \times 1$ et on écrira indifféremment $x = (x_1, \dots, x_d)$ ou $x = (x_1 \dots x_d)^{\mathsf{T}}$. Pour $x = (x_1 \dots x_d)^{\mathsf{T}}$ et $y = (y_1 \dots y_d)^{\mathsf{T}}$, on a $x^{\mathsf{T}}y = x_1y_1 + \dots + x_dy_d =< x, y >$ et xy^{T} est la matrice de terme général x_iy_j.

Pour $X \in L^2_d$, on définit:

$$K(X) = \mathbb{E}[(X - \mathbb{E}(X))(X - \mathbb{E}(X))^{\mathsf{T}}] = \mathbb{E}(XX^{\mathsf{T}}) - \mathbb{E}(X)[\mathbb{E}(X)]^{\mathsf{T}}. \tag{4.20}$$

$K(X)$ s'appelle la matrice de covariance ou la matrice de dispersion de X. On a

$$K(X) = \begin{pmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) & \dots & \dots & \dots & \mathrm{Cov}(X_1, X_d) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Var}(X_2) & \dots & \dots & \dots & \mathrm{Cov}(X_2, X_d) \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \mathrm{Cov}(X_d, X_1) & \dots & \dots & \dots & \dots & \mathrm{Var}(X_d) \end{pmatrix}.$$

Noter que, si les composantes $X_1, \dots, X_d$ sont indépendantes, $K(X)$ est diagonale.

Proposition 4.5.4. *Soit $X \in L^2_d$. On a*
(i) $K(\alpha X) = \alpha^2 K(X)$, $\alpha \in \mathbb{R}$; $K(X + a) = K(X)$, $a \in \mathbb{R}^d$; $K^{\mathsf{T}}(X) = K(X)$.
(ii) Pour tout $\lambda \in \mathbb{R}^d$, $\lambda^{\mathsf{T}} K(X)\lambda \geq 0$.
(iii) Soit M une matrice déterministe $r \times d$, on a $K(MX) = MK(X)M^{\mathsf{T}}$.

Preuve: (i) résulte de la définition (4.20).

(ii) Vu (i), on peut supposer $\mathbb{E}(X) = 0$. Alors

$$\lambda^{\mathsf{T}} K(X)\lambda = \lambda^{\mathsf{T}} \mathbb{E}(XX^{\mathsf{T}})\lambda = \mathbb{E}(\lambda^{\mathsf{T}} XX^{\mathsf{T}}\lambda) = \mathbb{E}|\lambda^{\mathsf{T}} X|^2 \geq 0.$$

(iii) Vu (i), on peut supposer $\mathbb{E}(X) = 0$. Alors

$$K(MX) = \mathbb{E}(MX(MX)^{\mathsf{T}}) = \mathbb{E}(MXX^{\mathsf{T}}M^{\mathsf{T}}) = M\mathbb{E}(XX^{\mathsf{T}})M^{\mathsf{T}} = MK(X)M^{\mathsf{T}}. \diamond$$

Les points (i) et (ii) montrent que $K(X)$ est symétrique semi-définie positive.

Théorème 4.5.5. *Soient $X, Y \in L^2_d$ des vecteurs aléatoires indépendants, on a $K(X+Y) = K(X)+K(Y)$. En particulier, si $d = 1$, $Var(X+Y) = Var(X)+ Var(Y)$ si les v.a.r. X et Y sont indépendantes.*

Preuve: On peut supposer $\mathbb{E}(X) = \mathbb{E}(Y) = 0$. Alors $K(X + Y) = \mathbb{E}((X + Y)(X + Y)^{\mathsf{T}}) = \mathbb{E}(XX^{\mathsf{T}}) + \mathbb{E}(YY^{\mathsf{T}})$ puisque, vu l'indépendance, $\mathbb{E}(XY^{\mathsf{T}}) = \mathbb{E}(X)\mathbb{E}(Y^{\mathsf{T}}) = 0$ et de même $\mathbb{E}(YX^{\mathsf{T}}) = 0$. $\diamond$

4.5.4. La matrice de dispersion donne des renseignements sur le support de la loi de X.

Proposition 4.5.6. *Soit* $X \in L^2_d$. *On a* $\mathbb{P}(X - \mathbb{E}(X) \in \operatorname{Im} K(X)) = 1$.

Preuve: Comme toujours on peut supposer $\mathbb{E}(X) = 0$. Soit $V = \operatorname{Im} K(X)$. Si $\dim(V) = d$, il n'y a rien à montrer. Supposons $\dim(V) = r < d$. Il existe $a_1, \ldots, a_{d-r} \in \operatorname{Ker}(X)$ tels que $x \in V$ ssi $a_k^{\mathsf{T}} x = 0$, $k = 1, \ldots, d-r$ (pour voir cela il suffit de se placer dans une base où $K(X)$ est diagonale). On a alors, vu la prop. 4.5.4,

$$\mathbb{E}(a_k^{\mathsf{T}} X)^2 = \operatorname{Var}(a_k^{\mathsf{T}} X) = K(a_k^{\mathsf{T}} X) = a_k^{\mathsf{T}} K(X) a_k = 0$$

d'où $a_k^{\mathsf{T}} X = 0$ p.s. et $X \in V$ p.s. $\diamond$

4.6. Calcul de lois

Soit X une v.a. à valeurs $\mathbb{R}^d$. Une probabilité μ sur $\mathbb{R}^d$ est la loi de X ssi, pour toute $f \in \mathcal{B}(\mathbb{R}^d)$, $\mathbb{E}(f(X)) = \int f \, d\mu$, soit encore, compte tenu de la prop. 3.5.4 et du cor. 3.5.6, ssi:

$$\text{pour toute } f \text{ positive de } C_k^\infty, \ \ \mathbb{E}(f(X)) = \int f \, d\mu. \tag{4.21}$$

4.6.1. Commençons par deux exemples élémentaires.

Exemple 1. Soit X une v.a.r. de densité (loi de Cauchy) $q(x) = \frac{1}{\pi(1+x^2)}$. On pose $Y = \mathrm{e}^X$. Quelle est la loi de Y ? Soit $f \in C_k^+$ arbitraire, on a, posant $y = \mathrm{e}^x$,

$$\mathbb{E}(f(Y)) = \mathbb{E}(f(\mathrm{e}^X)) = \int_{-\infty}^{+\infty} f(\mathrm{e}^x) \frac{dx}{\pi(1+x^2)} = \int_0^{+\infty} f(y) \frac{dy}{\pi y(1 + (\log y)^2)}.$$

Donc (4.21) Y a pour densité $\frac{1}{\pi y(1+(\log y)^2)} 1_{\mathbb{R}^+}(y)$.

Exemple 2. Soit X une v.a.r. de densité $N_1(0,1)$. On pose $Z = X^2$. Quelle est la loi de Z ? De même, pour $f \in C_k^+$ arbitraire,

$$\mathbb{E}(f(Z)) = \mathbb{E}(f(X^2)) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(x^2) \mathrm{e}^{-x^2/2} \, dx.$$

L'application $x \mapsto x^2$ n'étant pas une bijection de $\mathbb{R}$ sur $\mathbb{R}^+$, on ne peut pas poser brutalement $z = x^2$, mais on a

$$\mathbb{E}(f(Z)) = \mathbb{E}(f(X^2)) = \frac{2}{\sqrt{2\pi}} \int_0^{+\infty} f(x^2) \mathrm{e}^{-x^2/2} \, dx = \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} f(z) \mathrm{e}^{-z/2} \frac{dz}{\sqrt{z}}.$$

Donc (4.21) Z a pour densité $\frac{1}{\sqrt{2\pi}} e^{-z/2} z^{-1/2} 1_{\mathbb{R}^+}(z)$ i.e. $Z \sim G(\frac{1}{2}, \frac{1}{2})$.

4.6.2. Rappelons la formule de changement de variables dans $\mathbb{R}^d$. Si ϕ est un difféomorphisme de l'ouvert U sur l'ouvert V, on a, pour toute $f \in \mathcal{B}^+(\mathbb{R}^d)$,

$$\int_V f(v) \, dv = \int_U f(\phi(u)) |J(\phi)(u)| \, du. \tag{4.22}$$

où $J(\phi)$ est le déterminant de la matrice des $\frac{\partial \phi_j}{\partial u_k}$. Rappelons également que $J(\phi)(u) = \{J(\phi^{-1})(\phi(u))\}^{-1}$. Il en résulte:

Proposition 4.6.1. *Soit X un vecteur aléatoire de densité h. On suppose que $X \in D$ p.s., D ouvert de $\mathbb{R}^d$. Soient ψ un difféomorphisme de D sur un ouvert Δ et $Y = \psi(X)$, alors Y a pour densité*

$$h(\psi^{-1}(y))|J(\psi^{-1})(y)|1_\Delta(y).$$

Preuve: On a, pour toute $f \in \mathcal{B}^+(\mathbb{R}^d)$,

$$\mathbb{E}(f(Y)) = \mathbb{E}(f(\psi(X))) = \int_D f(\psi(x))h(x)\,dx = \int_\Delta f(y)h(\psi^{-1}(y))|J(\psi^{-1})(y)|\,dy. \diamond$$

Une première conséquence de (4.22) est la suivante (voir aussi 5.1.1):

Proposition 4.6.2. *Soient X et Y deux v.a. à valeurs $\mathbb{R}^d$, indépendantes, de densité respectives f et g. Alors la v.a. $S = X + Y$ a pour densité $h = f * g$ définie par*

$$h(u) = \int f(v)g(u-v)\,dv.$$

Preuve: On a, pour toute $\phi \in C_k^+$,

$$\mathbb{E}(\phi(S)) = \int\int \phi(x+y)f(x)g(y)\,dxdy = \int\int \phi(u)f(v)g(u-v)\,dudv = \int \phi(u)h(u)\,du.$$

Application. Soient X et Y des v.a.r. indépendantes de même loi la loi uniforme sur $[0,1]$. Quelle est la loi de $S = X + Y$? Soit h la densité de S. On a (attention aux fonctions indicatrices):

$$h(u) = \int 1_{[0,1]}(v)1_{[0,1]}(u-v)\,dv = \int_0^1 1_{[0,1]}(u-v)\,dv = \int_0^1 1_{[u-1,u]}(v)\,dv.$$

Si $0 \le u \le 1$, $h(u) = \int_0^u dv = u$, si $1 \le u \le 2$, $h(u) = \int_{u-1}^1 dv = 2-u$ et évidemment $h(u) = 0$ si $u \notin [0,2]$.

4.6.3. Exemple 3. Soient X et Y des v.a.r. indépendantes de lois respectives $G(a,c)$ et $G(b,c)$ (4.12), $a,b,c > 0$. On pose $S = X + Y$, $T = \frac{X}{X+Y}$. On veut calculer la loi du couple (S,T). Vu l'indépendance, le couple (X,Y) a pour densité

$$h_{X,Y}(x,y) = \frac{c^{a+b}}{\Gamma(a)\Gamma(b)}\mathrm{e}^{-c(x+y)}x^{a-1}y^{b-1}1_{]0,+\infty[}(x)1_{]0,+\infty[}(y).$$

Soit ϕ l'application $(x,y) \mapsto (s = x+y,\ t = \frac{x}{x+y})$. ϕ est un difféomorphisme de $]0,+\infty[\times]0,+\infty[$ sur $]0,+\infty[\times]0,1[$. De plus $J(\phi^{-1})(s,t) = -s$. La densité de (S,T) est donc (prop.4.6.1)

$$h_{S,T}(s,t) = \frac{c^{a+b}}{\Gamma(a)\Gamma(b)}\mathrm{e}^{-cs}s^{a+b-1}t^{a-1}(1-t)^{b-1}1_{]0,+\infty[}(s)1_{]0,1[}(t).$$

Le cor.4.5.3 montre que S et T sont indépendantes, que S a pour densité

$$h_S(s) = \frac{c^{a+b}}{\Gamma(a+b)}\, \mathrm{e}^{-cs} s^{a+b-1} 1_{]0,+\infty[}(s)$$

i.e. $S \sim G(a+b, c)$ et que T a pour densité

$$h_T(t) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\, t^{a-1}(1-t)^{b-1} 1_{]0,1[}(t).$$

Puisque h_T est une densité de probabilité, on a montré au passage la formule

$$\int_0^1 t^{a-1}(1-t)^{b-1}\, dt = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}. \tag{4.23}$$

4.6.4. L'exemple suivant sera très utile pour simuler des v.a.r. gaussiennes.

Proposition 4.6.3. *Soient (X,Y) un couple de v.a.r. indépendantes de même loi $U(0,1)$. On pose $U = \sqrt{-2\log X}.\cos(2\pi Y)$, $V = \sqrt{-2\log X}.\sin(2\pi Y)$. Alors les v.a. U et V sont indépendantes de même loi $N_1(0,1)$.*

Preuve: Soit $\psi\,:\,(x,y) \mapsto (u = \sqrt{-2\log x}.\cos(2\pi y), v = \sqrt{-2\log x}.\sin(2\pi y)$. ψ est un difféomorphisme de $D =]0,1[\times]0,1[$ sur $\Delta = \mathbb{R}^2 \setminus (\mathbb{R}^+ \times \{0\})$. On a $J(\psi)(x,y) = -\frac{2\pi}{x}$, et, vu que $u^2+v^2 = -2\log x$, $J(\psi^{-1})(u,v) = -\frac{1}{2\pi}\mathrm{e}^{-(u^2+v^2)/2}$. Le couple (X,Y) a pour densité $1_D(x,y)$. Donc (prop. 4.6.1) $(U,V) = \psi(X,Y)$ a pour densité

$$\frac{1}{2\pi}\mathrm{e}^{-(u^2+v^2)/2} 1_\Delta(u,v) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-u^2/2}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-v^2/2} \text{ p.p. } \diamond$$

4.6.5. Exemple 4. Soit (X,Y) un couple de v.a.r. indépendantes de même loi $N_1(0,1)$. On pose $T = \frac{Y}{X}$ (noter que $\mathbb{P}(X=0)=0$). Quelle est la loi de T ? Evidemment on ne peut pas appliquer directement la prop. 4.6.1. On choisit d'abord une v.a. $U = f(X,Y)$ telle qu'on puisse utiliser la prop. 4.6.1 pour obtenir la densit de (T,U) puis on obtient la loi de T comme marginale. Ici on peut choisir $U = X$.

Soit $\psi\,:\,(x,y) \mapsto (t = y/x, u = x)$. Alors ψ est un difféomorphisme de $D = \mathbb{R}\times\mathbb{R}^*$ sur $\Delta = \mathbb{R}\times\mathbb{R}^*$. On a $x = u, y = tu$, et $J(\psi^{-1})(u,v) = -u$. Le couple (X,Y) a pour densité $\frac{1}{2\pi}\mathrm{e}^{-(x^2+y^2)/2}1_D(x,y)$. Alors (prop. 4.6.1) $(T,U) = \psi(X,Y)$ a pour densité $\frac{1}{2\pi}\mathrm{e}^{-u^2(1+t^2)/2}|u|1_\Delta(t,u) = \frac{1}{2\pi}\mathrm{e}^{-u^2(1+t^2)/2}|u|$ p.p.

Donc T a pour densité

$$q(t) = \int_{-\infty}^{+\infty} \frac{1}{2\pi}\mathrm{e}^{-u^2(1+t^2)/2}|u|\, du = \frac{1}{\pi}\int_0^{+\infty} \mathrm{e}^{-u^2(1+t^2)/2} u\, du = \frac{1}{\pi(1+t^2)}.$$

La v.a. T suit donc une loi de Cauchy.

En fait, il est souvent plus rapide de calculer directement $\mathbb{E}(f(T))$. Ici, par exemple, passant en coordonnées polaires, on a:

$$\mathbb{E}(f(T)) = \frac{1}{2\pi}\int\int f(\frac{y}{x})\mathrm{e}^{-\frac{1}{2}(x^2+y^2)})\, dxdy = \frac{1}{2\pi}\int_0^{2\pi}\int_0^{\infty} f(\tan\theta)\mathrm{e}^{-\frac{\rho^2}{2}}\rho\, d\theta d\rho$$

$$= \frac{1}{\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} f(\tan\theta)\, d\theta = \frac{1}{\pi} \int_{-\infty}^{+\infty} f(t) \frac{1}{1+t^2}\, dz.$$

4.6.6. Exemple 5. Soit (X, Y) un couple de v.a.r. indépendantes de même loi $N_1(0, 1)$. On pose $U = X, V = X^2 + Y^2$. Quelle est la loi du couple (U, V) ? L'application $(x, y) \mapsto (x, x^2 + y^2)$ n'étant pas une bijection, on ne peut utiliser la prop. 4.6.1. Soit $f \in C_k^+(\mathbb{R}^2)$ arbitraire. On a

$$\mathbb{E}(f(U,V)) = \mathbb{E}(f(X, X^2 + Y^2)) = \frac{1}{2\pi} \int_{\mathbb{R}^2} f(x, x^2 + y^2) \mathrm{e}^{-(x^2+y^2)/2}\, dxdy$$
$$= \frac{1}{2\pi} \int_{\mathbb{R}\times\mathbb{R}^+} \ldots + \frac{1}{2\pi} \int_{\mathbb{R}\times\mathbb{R}^-} \ldots.$$

Considérons l'application $(x, y) \mapsto (u = x, v = x^2+y^2)$. C'est d'une part une bijection de $\mathbb{R} \times \mathbb{R}^+$ sur $\Gamma = \{(u, v),\ v \geq u^2\}$ et alors $x = u, y = \sqrt{v - u^2}$ et d'autre part une bijection de $\mathbb{R} \times \mathbb{R}^-$ sur Γ et dans ce cas $x = u, y = -\sqrt{v - u^2}$. Dans les deux cas, $|J| = \frac{1}{2\sqrt{v-u^2}}$. On obtient

$$\mathbb{E}(f(U,V)) = \frac{1}{2\pi} \int_{\Gamma} f(u, v) \frac{\mathrm{e}^{-v/2}}{\sqrt{v - u^2}}\, dudv.$$

Le couple a donc pour densité $\frac{\mathrm{e}^{-v/2}}{2\pi\sqrt{v-u^2}} 1_\Gamma(u, v)$.

4.6.7. Exemple 6. On ne rencontre pas toujours des v.a. ayant une densité par rapport à la mesure de Lebesgue. Soit X une v.a.r. de densité $\mathrm{e}^{-x} 1_{\mathbb{R}^+}(x)$. On pose $U = [X], V = X - [X]$ où $[x]$ désigne la partie entière de x. Quelle est la loi de (U, V) ? Quelles sont les lois de U et de V ? Les v.a. U et V sont-elles indépendantes ?

Soit $f \in C_k^+(\mathbb{R}^2)$ arbitraire. On a

$$\mathbb{E}(f(U,V)) = \int_0^{+\infty} f([x], (x - [x])\mathrm{e}^{-x}\, dx$$
$$= \sum_{k=0}^{\infty} \int_k^{k+1} f(k, x - k)\mathrm{e}^{-x}\, dx = \sum_{k=0}^{\infty} \int_0^1 f(k, t)\mathrm{e}^{-k}\mathrm{e}^{-t}\, dt.$$

Si on note ν la mesure sur $\mathbb{N}$ définie par $\nu(\{k\}) = 1$ et λ la mesure de Lebesgue sur $[0, 1]$, ce calcul implique que la loi de (U, V) est la probabilité $\mathrm{e}^{-k}\mathrm{e}^{-t}.\nu \otimes \lambda$.

Prenant $f(u, v) = \phi(u)$, on a

$$\mathbb{E}(\phi(U)) = \sum_{k=0}^{\infty} \phi(k)\mathrm{e}^{-k}(1 - \mathrm{e}^{-1}) = \sum_{k=0}^{\infty} \phi(k)(\mathrm{e}^{-1})^k(1 - \mathrm{e}^{-1})$$

et U suit une loi géométrique de paramètre e^{-1}.

Prenant $f(u, v) = \psi(u)$, on a

$$\mathbb{E}(\psi(V)) = \int_0^1 \sum_{k=0}^{\infty} \mathrm{e}^{-k}\psi(t)\mathrm{e}^{-t}\, dt = \int_0^1 (1 - \mathrm{e}^{-1})^{-1}\psi(t)\mathrm{e}^{-t}\, dt$$

et V a pour densité $\frac{\mathrm{e}}{\mathrm{e}-1}\mathrm{e}^{-t}1_{]0,1[}(t)$.

Enfin $\mathbb{E}(\phi(U)\psi(V)) = \mathbb{E}(\phi(U))\,\mathbb{E}(\psi(V))$ et U et V sont indépendantes (th. 4.4.4).

4.6.8. Loi des min et des max. Soient $X_1, X_2, \dots, X_n$ des v.a. réelles indépendantes de fonction de répartition $F_1, F_2, \dots, F_n$. On pose

$$U = \min_{1\le k\le n} X_k, \quad V = \max_{1\le k\le n} X_k.$$

D'une part

$$\mathbb{P}(V \le t) = \mathbb{P}(X_1 \le t, \dots, X_n \le t) = \prod_{k=1}^{n} \mathbb{P}(X_k \le t) = \prod_{k=1}^{n} F_k(t)$$

et V a pour fonction de répartition $F_V(t) = \prod_{k=1}^{n} F_k(t)$. D'autre part

$$\mathbb{P}(U > t) = \mathbb{P}(X_1 > t, \dots, X_n > t) = \prod_{k=1}^{n} \mathbb{P}(X_k > t) = \prod_{k=1}^{n}(1 - F_k(t))$$

et U a pour fonction de répartition $F_U(t) = 1 - \prod_{k=1}^{n}(1 - F_k(t))$.

Si les X_k ont même loi, pour tout k, $F_k(t) = F(t)$ et

$$F_V(t) = (F(t))^n, \quad F_U(t) = 1 - (1 - F(t))^n.$$

Si, de plus, les X_k ont une densité, F est dérivable et on obtient les densités de U et V en dérivant $F_U(t)$ et $F_V(t)$.

4.7. Conditionnement

4.7.1. Soient A un événement tel que $\mathbb{P}(A) > 0$ et Y une v.a à valeurs $\mathbb{R}^d$. Posons, pour $\Gamma \in \mathcal{B}(\mathbb{R}^d)$,

$$\mu_Y(\Gamma|A) = \mathbb{P}(Y \in \Gamma|A) = \frac{1}{\mathbb{P}(A)}\,\mathbb{P}(A \cap \{Y \in \Gamma\}). \tag{4.24}$$

Alors, A étant fixé, $\Gamma \mapsto \mu_Y(\Gamma|A)$ est une probabilité sur $\mathbb{R}^d$ qu'on appelle loi conditionnelle de Y sachant A. De même, pour $\phi \in L^1(\mu_Y)$,

$$\int \phi(y)\, d\mu_Y(y|A) = \mathbb{E}(\phi(Y)|A) = \frac{1}{\mathbb{P}(A)} \int_A \phi(Y)\, d\mathbb{P} \tag{4.25}$$

s'appelle l'espérance conditionnelle de $\phi(Y)$ sachant A.

4.7.2. Considérons une v.a. à valeurs E fini ou dénombrable telle que, pour tout $a \in E$, $\mathbb{P}(X = a) > 0$ et Y une v.a à valeurs $\mathbb{R}^d$. Prenant $A = \{X = a\}$, on obtient la loi conditionnelle de Y sachant que $X = a$ définie par

$$\mu_Y(\Gamma|X = a) = \mathbb{P}(Y \in \Gamma|X = a) = \frac{1}{\mathbb{P}(X = a)}\,\mathbb{P}(X = a, Y \in \Gamma) \tag{4.26}$$

et, pour $\phi \in L^1(\mu_Y)$, l'espérance conditionnelle de $\phi(Y)$ sachant que $X = a$ définie par

$$\mathbb{E}(\phi(Y)|X = a) = \frac{1}{\mathbb{P}(X = a)} \int_{\{X=a\}} \phi(Y)\, d\mathbb{P}. \tag{4.27}$$

4.7.3. Considérons maintenant une v.a. X à valeurs $\mathbb{R}^p$ de densité $q(x)$ et Y une v.a à valeurs $\mathbb{R}^d$. Les formules (4.26) et (4.27) n'ont plus de sens puisque, pour tout a, $\mathbb{P}(X = a) = 0$. Supposons que (X, Y) ait une densité continue $h(x, y)$ et que $q(x) = \int h(x, y)\, dy > 0$. Soient $B(a, \delta)$ la boule dans $\mathbb{R}^p$ de centre a et de rayon δ et $|B(a, \delta)|$ son volume. On a, lorsque $\delta \to 0$,

$$\begin{aligned}\mathbb{P}(Y \in \Gamma|X \in B(a, \delta)) &= \frac{\mathbb{P}(X \in B(a, \delta), Y \in \Gamma)}{\mathbb{P}(X \in B(a, \delta))} = \frac{\int_{B(a,\delta)\times\Gamma} h(x, y)\, dxdy}{\int_{B(a,\delta)} q(x)\, dx} \\ &= \int_\Gamma \frac{|B(a, \delta)|^{-1} \int_{B(a,\delta)} h(x, y)\, dx}{|B(a, \delta)|^{-1} \int_{B(a,\delta)} q(x)\, dx}\, dy \to \int_\Gamma \frac{h(a, y)}{q(a)}\, dy.\end{aligned}$$

Il est donc naturel d'appeler loi conditionnelle de Y sachant que $X = a$ la loi de densité $h(a, y)/q(a)$. Ceci conduit à:

Définition 4.7.1. *Soient (X, Y) un couple de v.a. à valeurs $\mathbb{R}^p \times \mathbb{R}^d$ de densité $h(x, y)$ et $q(x) = \int h(x, y)\, dy$ la densité de X. On appelle densité conditionnelle de Y sachant que $X = x$ la fonction*

$$h(y\,|\,x) = \frac{h(x, y)}{q(x)} \text{ si } q(x) > 0, \ = \text{densité arbitraire si } q(x) = 0.$$

Remarque 1. Noter que $\mathbb{P}(X \in \{q = 0\}) = \int_{\{q=0\}} q(x)\, dx = 0$.

Remarque 2. On voit donc que $h(y\,|\,x)$ est le quotient de la densité de (X, Y) par la densité de X. C'est tout simplement l'analogue de la formule, pour des v.a. entières, $\mathbb{P}(Y = n\,|\,X = p) = \mathbb{P}(X = p, Y = n) \,/\, \mathbb{P}(X = p)$.

La loi de densité $h(y\,|\,x)$ s'appelle la loi conditionnelle de Y sachant que $X = x$ et, pour $\phi \in L^1(\mu_Y)$,

$$\mathbb{E}(\phi(Y)|X = x) := \int \phi(y) h(y\,|\,x)\, dy$$

s'appelle l'espérance conditionnelle de $\phi(Y)$ sachant que $X = x$. Si $d = 1$, on peut choisir $\phi(y) = y$, on obtient l'espérance conditionnelle de Y sachant que $X = x$. L'énoncé suivant est à comparer au lem. 4.2.6.

Proposition 4.7.2. *Soit (X, Y) un couple de v.a. à valeurs $\mathbb{R}^p \times \mathbb{R}$ de densité $h(x, y)$ avec $Y \in L^2$. Alors:*

$$\inf\{\, \mathbb{E}[(Y - f(X))^2],\, f \in L^2(\mu_X)\,\} = \mathbb{E}[(Y - \hat{f}(X))^2] \text{ où } \hat{f}(x) = \mathbb{E}(Y|X = x).$$

Preuve: Pour toute $g \in L^2(\mu_X)$, on a $\mathbb{E}((Y - \hat{f}(X))g(X)) = 0$. En effet, sur $\{q(x) > 0\}$, $\hat{f}(x) = \frac{1}{q(x)} \int yh(x,y)\,dy$ et, vu la remarque 1,

$$\mathbb{E}(Yg(X)) = \int_{\{q>0\}} yg(x)h(x,y)\,dxdy = \int_{\{q>0\}} g(x)q(x)\frac{1}{q(x)} \int yh(x,y)\,dy\,dx$$
$$= \int_{\{q>0\}} g(x)\hat{f}(x)q(x)\,dx = \mathbb{E}(g(X)\hat{f}(X)).$$

On en déduit:

$$\mathbb{E}[(Y - f(X))^2] = \mathbb{E}[(Y - \hat{f}(X) + \hat{f}(X) - f(X))^2]$$
$$= \mathbb{E}[(Y - \hat{f}(X))^2] + \mathbb{E}[(\hat{f}(X) - f(X))^2] + 2\mathbb{E}[(Y - \hat{f}(X))(\hat{f}(X) - f(X))]$$
$$= \mathbb{E}[(Y - \hat{f}(X))^2] + \mathbb{E}[(\hat{f}(X) - f(X))^2]$$

et le résultat cherché. $\diamond$

Exemple. Soient Y, Z des v.a.r. indépendantes de même densité $\lambda e^{-\lambda y} 1_{\mathbb{R}^+}(y)$. On pose $X = Y + Z$. On veut calculer la loi conditionnelle de Y sachant que $X = x$ et $\mathbb{E}(Y|X = x)$.

Pour appliquer la def.4.7.1, il faut calculer la densité du couple (X, Y). On a

$$\mathbb{E}(\phi(X,Y)) = \lambda^2 \int_0^\infty \int_0^\infty \phi(y+z)e^{-\lambda(y+z)}\,dydz = \lambda^2 \int_0^\infty \int_0^x \phi(x,y)e^{-\lambda x}\,dxdy$$

et (X, Y) a pour densité $h(x,y) = \lambda^2 e^{-\lambda x} 1_{\{0 \le y \le x\}}$. La densité de X est alors

$$q(x) = \lambda^2 e^{-\lambda x} \int_0^x dy = \lambda^2 x e^{-x} \text{ si } x > 0; \;\; q(x) = 0 \text{ si } x \le 0.$$

Finalement, pour $x > 0$ (noter que $\mathbb{P}(X \le 0) = 0$),

$$h(y \,|\, x) = \frac{h(x,y)}{q(x)} = \frac{1}{x} 1_{[0,x]}(y).$$

La la loi conditionnelle de Y sachant que $X = x$ est donc la loi uniforme sur $[0, x]$ et

$$\mathbb{E}(Y|X = x) = \int yh(y \,|\, x)\,dy = \frac{1}{x} \int_0^x y\,dy = \frac{x}{2}$$

qui est évidemment la moyenne de la loi $U(0, x)$.

4.8. Simulation

Soit μ une probabilité sur $\mathbb{R}^d$. Simuler la loi μ, c'est construire une suite $x_1, x_2, \ldots, x_n, \ldots$ de points de $\mathbb{R}^d$ censés être le résultat de tirages indépendants de points de $\mathbb{R}^d$ selon la loi μ i.e. les valeurs prises par une suite $X_1, X_2, \ldots, X_n, \ldots$ de v.a. indépendantes de loi μ.

4.8.1. Nombres au hasard. En général, la fonction “random” d'un ordinateur fournit une suite de nombres entre 0 et 1 censés être le résultat de tirages indépendants selon la loi uniforme sur $[0,1]$. Ces nombres sont obtenus par un algorithme qui fournit des nombres ayant les mêmes propriétés qu'une suite de tirages indépendants selon $U(0,1)$. A ce sujet, voir la sous-section 6.4.2. Le problème est donc de construire à partir d'une suite $U_1, U_2, \ldots, U_n, \ldots$ de v.a. indépendantes de loi $U(0,1)$ une suite $X_1, X_2, \ldots, X_n, \ldots$ de v.a. indépendantes de loi μ.

4.8.2. Simulation de v.a. réelles. Soit μ une probabilité sur $\mathbb{R}$ de fonction de répartition F. On pose $F^{-1}(u) = \inf(t,\, F(t) \geq u)$. On sait (prop. 4.3.2) que, si $U \sim U(0,1)$, $F^{-1}(U)$ a pour loi μ. Donc, si $(U_n,\, n \geq 1)$ est une suite de v.a. indépendantes de loi $U(0,1)$, $(F^{-1}(U_n),\, n \geq 1)$ est une suite de v.a. indépendantes de loi μ.

Exemple. Soit $(p_k,\, k = 0, \ldots, n)$ une probabilité sur $\{0, 1, \ldots, n\}$. Soit $F(t)$ sa fonction de répartition. On pose

$$a_0 = 0,\ a_1 = p_0,\ a_2 = p_0 + p_1, \ldots,\ a_n = p_0 + \ldots + p_{n-1},\ a_{n+1} = 1.$$

On a

$$F(t) = 0 = a_0 \text{ si } t < 0,\ F(t) = a_1 \text{ si } 0 \leq t < 1,\ F(t) = a_2 \text{ si } 2 \leq t < 3, \ldots$$

et

$$F^{-1}(u) = k \text{ si } a_k \leq u < a_{k+1},\quad k = 0, 1, \ldots, n.$$

Si $\mu = f.\lambda$, $F(t) = \int_{-\infty}^{t} f(x)\, dx$. Il n'est pas toujours (en fait pas souvent) possible de calculer F et F^{-1}. C'est en particulier le cas pour la loi $N_1(0,1)$.

4.8.3. Simulation de v.a. gaussiennes réelles. Soit $(U_n,\, n \geq 1)$ une suite de v.a. indépendantes de loi $U(0,1)$, on pose, pour $n \geq 1$,

$$X_{2n-1} = \sqrt{-2 \log U_{2n-1}}.\cos(2\pi U_{2n}),\quad X_{2n} = \sqrt{-2 \log U_{2n-1}}.\sin(2\pi U_{2n}).$$

Alors d'après la prop. 4.6.3, $(X_n,\, n \geq 1)$ est une suite de v.a. indépendantes de loi $N_1(0,1)$. Pour simuler la loi $N_1(m, \sigma^2)$, il suffit de remarquer que, si $Y \sim N_1(0,1)$, alors $X = m + \sigma Y \sim N_1(m, \sigma^2)$.

4.8.4. La méthode de rejet. Soient $(Z_n,\, n \geq 1)$ une suite de v.a. à valeurs $(E, \mathcal{E})$ et $B \in \mathcal{E}$. On considère $\nu = \inf(n \geq 1,\, Z_n \in B)$ avec la convention $\inf \emptyset = +\infty$. Alors $\nu(\omega)$ est la premier n tel que $Z_n(\omega) \in B$ et si, pour tout n, $Z_n(\omega) \notin B$, $\nu(\omega) = +\infty$. ν est donc une v.a. à valeurs $\overline{\mathbb{N}}$. Si $\mathbb{P}(\nu < +\infty) = 1$, on peut définir une v.a. Z_ν par $Z_\nu(\omega) = Z_n(\omega)$ sur $\{\omega,\, \nu(\omega) = n\}$. La méthode de rejet repose sur:

Proposition 4.8.1. *Soient $(Z_n,\, n \geq 1)$ une suite de v.a. indépendantes à valeurs $(E, \mathcal{E})$ de même loi μ et $B \in \mathcal{E}$ avec $\mu(B) > 0$. On pose $\nu_1 = \inf(n \geq 1,\, Z_n \in B)$, $\ldots$, $\nu_r = \inf(n > \nu_{r-1},\, Z_n \in B)$, $\ldots$. Alors, pour tout $r \geq 1$, $\mathbb{P}(\nu_r < +\infty) = 1$ et $(Z_{\nu_r},\, r \geq 1)$ est une suite de v.a. indépendantes de loi ρ donnée par*

$$\rho(A) = \frac{\mu(A \cap B)}{\mu(B)} = \mathbb{P}(Z_1 \in A \mid Z_1 \in B)$$

i.e. ρ est donc la loi conditionnelle de Z_1 sachant que $Z_1 \in B$.

Preuve: Notons d'abord que

$$\mathbb{P}(\nu_1 = k) = \mathbb{P}(Z_1 \notin B, \ldots, Z_{k-1} \notin B, Z_k \in B) = (1 - \mu(B))^{k-1}\mu(B) \tag{4.28}$$

d'où $\mathbb{P}(\nu_1 < +\infty) = \sum_{k\geq 1} \mathbb{P}(\nu_1 = k) = 1$. Supposons que $\mathbb{P}(\nu_{r-1} < +\infty) = 1$, alors

$$\mathbb{P}(\nu_r < +\infty) = \sum_{k\geq 1} \mathbb{P}(\nu_{r-1} = k, \nu_r < +\infty) = \sum_{j,k\geq 1} \mathbb{P}(\nu_{r-1} = k, \nu_r = k + j)$$
$$= \sum_{j,k\geq 1} \mathbb{P}(\nu_{r-1} = k, Z_{k+1} \notin B, \ldots, Z_{k+j-1} \notin B, Z_{k+j} \in B)$$
$$= \sum_{k\geq 1} \mathbb{P}(\nu_{r-1} = k) \sum_{j\geq 1} (1 - \mu(B))^{k-1}\mu(B) = \sum_{k\geq 1} \mathbb{P}(\nu_{r-1} = k) = \mathbb{P}(\nu_{r-1} < +\infty) = 1.$$

De même

$$\mathbb{P}(Z_{\nu_1} \in A) = \sum_{k\geq 1} \mathbb{P}(\nu_1 = k, Z_k \in A \cap B)$$
$$= \sum_{k\geq 1} \mathbb{P}(Z_1 \notin B, \ldots, Z_{k-1} \notin B, Z_k \in A\cap B) = \sum_{k\geq 1} (1-\mu(B))^{k-1}\mu(A\cap B) = \frac{\mu(A \cap B)}{\mu(B)}.$$

Supposons que $\mathbb{P}(Z_{\nu_1} \in A_1, \ldots, Z_{\nu_{r-1}} \in A_{r-1}) = \frac{\mu(A_1\cap B)}{\mu(B)} \ldots \frac{\mu(A_{r-1}\cap B)}{\mu(B)}$, alors

$$\mathbb{P}(Z_{\nu_1} \in A_1, \ldots, Z_{\nu_{r-1}} \in A_{r-1}, Z_{\nu_r} \in A_r)$$
$$= \sum_{j,k\geq 1} \mathbb{P}(Z_{\nu_1} \in A_1, \ldots, Z_{\nu_{r-1}} \in A_{r-1}, \nu_{r-1} = k, Z_{k+1} \notin B, \ldots, Z_{k+j-1} \notin B, Z_{k+j} \in A_r\cap B)$$
$$= \sum_{k\geq 1} \mathbb{P}(Z_{\nu_1} \in A_1, \ldots, Z_{\nu_{r-1}} \in A_{r-1}, \nu_{r-1} = k) \sum_{j\geq 1} (1 - \mu(B))^{j-1}\mu(A_r \cap B)$$
$$= \mathbb{P}(Z_{\nu_1} \in A_1, \ldots, Z_{\nu_{r-1}} \in A_{r-1}) \frac{\mu(A_r \cap B)}{\mu(B)} = \prod_{i=1}^{r} \frac{\mu(A_i \cap B)}{\mu(B)},$$

ce qui montre que les v.a. $(Z_{\nu_k},\ k \geq 1)$ sont indépendantes et de même loi. $\diamond$

En pratique, soit $z_1, \ldots, z_n, \ldots$ une suite de tirages indépendants selon la loi μ. On considère z_1. Si $z_1 \in B$, on pose $x_1 = z_1, k_1 = 1$. Sinon, on considère z_2. Si $z_2 \in B$, on pose $x_1 = z_2, k_1 = 2$. Sinon, on considère z_3. Si $z_3 \in B$, on pose $x_1 = z_3, k_1 = 3$. ... On construit ainsi x_1, k_1. On considère alors z_{k_1+1}. Si $z_{k_1+1} \in B$, on pose $x_2 = z_{k_1+1}, k_2 = k_1 + 1$. Sinon, on considère z_{k_1+2}. Si $z_{k_1+2} \in B$, on pose $x_2 = z_{k_1+2}, k_2 = k_1 + 2$. Sinon, on considère z_{k_1+3}. Si $z_{k_1+3} \in B$, on pose $x_2 = z_{k_1+3}, k_2 = k_1 + 3$. ... On construit ainsi x_2, k_2. On continue.... et on obtient une suite $x_1, \ldots, x_n, \ldots$ de tirages indépendants selon la loi $\nu(A) = \frac{\mu(A\cap B)}{\mu(B)}$.

Remarque 1. Vu (4.28), la v.a. $\nu_1 - 1$ suit une loi géométrique de paramètre $1-\mu(B)$ et on a $\mathbb{E}(\nu_1) = \frac{1}{\mu(B)}$. Il est intuitif (et facile à vérifier) que les v.a. $\nu_1, \nu_2-\nu_1, \ldots, \nu_r-\nu_{r-1}$

sont indépendantes et de même loi. On a donc $\mathbb{E}(\nu_1) = \mathbb{E}(\nu_2 - \nu_1) = \ldots = \mathbb{E}(\nu_r - \nu_{r-1}) = \frac{1}{\mu(B)}$. Donc, si $\mu(B)$ est très petit, cette simulation risque de prendre du temps.

4.8.5. Simulation de la loi uniforme sur un domaine de mesure de Lebesgue finie. Soit D un domaine de $\mathbb{R}^d$ tel que $\lambda(D) < +\infty$, λ étant la mesure de Lebesgue sur $\mathbb{R}^d$. On appelle loi uniforme sur D, la probabilité de densité $(\lambda(D))^{-1}1_D$. La prop. 4.8.1 donne immédiatement:

Corollaire 4.8.2. *Soient $D \subset \Delta$ deux domaines de $\mathbb{R}^d$ avec $\lambda(\Delta) < +\infty$ et $(Z_n,\, n \geq 1)$ une suite de v.a. indépendantes de loi la loi uniforme sur Δ. On pose $\nu_1 = \inf(n \geq 1,\, Z_n \in D)$, $\ldots$, $\nu_r = \inf(n > \nu_{r-1},\, Z_n \in D)$, $\ldots$. Alors, pour tout $r \geq 1$, $\mathbb{P}(\nu_r < +\infty) = 1$ et $(Z_{\nu_r},\, r \geq 1)$ est une suite de v.a. indépendantes de loi la loi uniforme sur D.*

Preuve: Il suffit de remarquer que, si μ est la loi uniforme sur Δ, la loi de X_{ν_1} est

$$\rho(A) = \frac{\mu(A \cap D)}{\mu(D)} = \frac{\lambda(A \cap D)}{\lambda(\Delta)} : \frac{\lambda(D)}{\lambda(\Delta)} = \frac{\lambda(A \cap D)}{\lambda(D)}$$

i.e. la loi uniforme sur D. $\diamond$

En pratique, si D est borné, on choisit $\Delta = [a_1, b_1] \times \ldots \times [a_d, b_d]$ et il est très facile de simuler la loi uniforme sur Δ et donc sur D.

4.8.5. Soit $D = \{(x, y),\ 0 \leq y < f(x)\} \subset \mathbb{R}^2$ où f est une densité de probabilité sur $\mathbb{R}$. Si (X, Y) est un couple de v.a. de loi uniforme sur D, alors X est une v.a.r. de densité f. Réciproquement, si X est une v.a.r. de densité f et si U est une v.a.r. de loi $U(0, 1)$, indépendante de X, alors $(X, Uf(X))$ suit la loi uniforme sur D et, plus généralement, $(X, aUf(X))$ $(a > 0)$ suit la loi uniforme sur $\Delta = \{(x, y),\ 0 \leq y < af(x)\}$. Ceci fournit une méthode, sachant simuler une loi de densité g, pour simuler une loi de densité f si $f \leq a\, g$ (nécessairement $a \geq 1$)). Plus précisemment:

Proposition 4.8.3. *Soient ρ une mesure σ-finie sur $(F, \mathcal{F})$ et $f, g \in \mathcal{F}^+$ telles que $\int f\, d\rho = \int g\, d\rho = 1$ et $f \leq ag$ ρ p.p. Soient $(Y_n,\, n \geq 1)$ une suite de v.a. indépendantes à valeurs $(F, \mathcal{F})$ de loi $g.\rho$ et $(U_n,\, n \geq 1)$ une suite de v.a.r. indépendantes de loi $U(0, 1)$ et indépendantes de $(Y_n,\, n \geq 1)$. On pose*

$$\nu_1 = \inf(n \geq 1,\, aU_n g(Y_n) < f(Y_n)\,), \ldots, \nu_r = \inf(n > \nu_{r-1},\, aU_n g(Y_n) < f(Y_n)\,), \ldots$$

Alors les v.a. $(Y_{\nu_r},\, r \geq 1)$ sont indépendantes de loi $f.\rho$.

Preuve: Soient $Z_n = (Y_n, U_n)$ et $\Gamma = \{(y, u),\ a.u.g(y) < f(y)\,\}$. On a alors $\nu_1 = \inf(n \geq 1,\, Z_n \in \Gamma)$,....

Lemme 4.8.4. *Pour toute $\phi \in \mathcal{F}^+$,*

$$\mathbb{E}(\phi(Y_1)1_{\{Z_1 \in \Gamma\}}) = \mathbb{E}(\phi(Y_1)1_{\{aU_1 g(Y_1) < f(Y_1)\}}) = \frac{1}{a} \int \phi(y) f(y)\, d\rho(y).$$

Preuve: Notons que $f1_{\{g=0\}} \leq ag1_{\{g=0\}} = 0$ ρ p.p. Alors

$$\mathbb{E}(\phi(Y_1)1_{\{Z_1\in\Gamma\}}) = \int\int_0^1 \phi(y)1_\Gamma(y,u)g(y)1_{\{g>0\}}(y)\,d\rho(y)du$$
$$= \int \phi(y)g(y)1_{\{g>0\}}(y)\int_0^1 1_{\{u<\frac{f(y)}{ag(y)}\}}\,du\,d\rho(y) = \int \phi(y)g(y)1_{\{g>0\}}(y)\frac{f(y)}{ag(y)}\,d\rho(y)$$
$$= \frac{1}{a}\int \phi(y)f(y)\,d\rho(y). \diamond$$

Prenant $\phi = 1$ dans le lem. 4.8.4, on obtient $\mathbb{P}(Z_1 \in \Gamma) = \frac{1}{a} > 0$ et on peut appliquer la prop. 4.8.1 . Les v.a. $(Z_{\nu_r},\, r \geq 1)$ (resp. $(Y_{\nu_r},\, r \geq 1)$) sont indépendantes de même loi que Z_{ν_1} (resp. Y_{ν_1}). Enfin on a (prop. 4.8.1 et lem. 4.8.4)

$$\mathbb{P}(Y_{\nu_1} \in A) = \frac{\mathbb{P}(Y_1 \in A, Z_1 \in \Gamma)}{\mathbb{P}(Z_1 \in \Gamma)} = \int_A f\,d\rho$$

et Y_{ν_1} a pour loi $f.\rho$. $\diamond$

Remarque 2. Vu que $P(Z_1 \in \Gamma) = \frac{1}{a}$, d'après la remarque 1, $\mathbb{E}(\nu_1) = \mathbb{E}(\nu_r - \nu_{r-1}) = a$. Si a est trop grand, cette méthode est coûteuse en temps.

4.9. Complément: échantillons ordonnés.

Dans cette section, on considère une probabilité μ sur $\mathbb{R}$. On note F sa fonction de répartition (def. 4.3.1). On rappelle que F est continue ssi $\mu(\{x\}) = 0$ pour tout $x \in \mathbb{R}$.

4.9.1. Echantillon ordonné. Soit $X_1, \ldots, X_n$ n v.a.r. indépendantes de loi μ. On appelle $X_1, \ldots, X_n$ un échantillon de taille n (ou n-échantillon) de la loi μ. Les $X_1, \ldots, X_n$ rangés par ordre croissant, qu'on note $X_{(1)}, \ldots, X_{(n)}$, s'appelle alors un échantillon ordonné de taille n de la loi μ. En particulier

$$X_{(1)} = \min_{1\leq i\leq n} X_i, \quad X_{(n)} = \max_{1\leq i\leq n} X_i.$$

Par exemple, si $X_1(\omega) = 4$, $X_2(\omega) = 5$, $X_3(\omega) = 1$, $X_4(\omega) = 2$, $X_5(\omega) = 4$, $X_6(\omega) = 4$, $X_7(\omega) = 2$, $X_8(\omega) = 3$, on a $X_{(1)}(\omega) = 1$, $X_{(2)}(\omega) = 2$, $X_{(3)}(\omega) = 2$, $X_{(4)}(\omega) = 3$, $X_{(5)}(\omega) = 4$, $X_{(6)}(\omega) = 4$, $X_{(7)}(\omega) = 4$, $X_{(8)}(\omega) = 5$.

Supposons F continue, on a alors, pour $i \neq j$,

$$\mathbb{P}(X_i = X_j) = \int\int 1_{\{x=y\}}\,d\mu(x)d\mu(y) = \int(\int 1_{\{y\}}(x)\,d\mu(x))\,d\mu(y) = 0,$$

et donc $\mathbb{P}(\cup_{i\neq j}\{X_i = X_j\})$ et $X_{(1)} < \ldots < X_{(n)}$ p.s.

Si on a un un échantillon ordonné de taille $2n+1$ de la loi μ, on pose

$$M_n = X_{(n+1)} \tag{4.29}$$

et M_n s'appelle la médiane de l'échantillon ou la médiane empirique.

4.9.2. Loi de $X_{(k)}$. Soit $X_1, \ldots, X_n$ un échantillon de taille n d!une loi μ. On pose

$$N_n^t = \sum_{i=1}^{n} 1_{]-\infty,t]}(X_i). \tag{4.30}$$

Alors $N_n^t \sim B(n, F(t))$ et $\{X_{(k)} \leq t\} = \{N_n^t \geq k\}$. On a donc, notant F_k la fonction de répartition de $X_{(k)}$,

$$\mathbb{P}(X_{(k)} \leq t) = \mathbb{P}(N_n^t \geq k) = \sum_{r=k}^{n} C_n^r (F(t))^r (1 - F(t))^{n-r}.$$

Vu que, pour $0 \leq \theta \leq 1$,

$$\frac{d}{d\theta} \sum_{r=k}^{n} \frac{n!}{r!(n-r)!} \theta^r (1-\theta)^{n-r} = \frac{n!}{(k-1)!(n-k)!} \theta^{k-1} (1-\theta)^{n-k}$$

(quand on dérive tous les termes se détruisent deux à deux sauf le premier), on obtient finalement:

Proposition 4.9.1. *Soient $X_1, \ldots, X_n$ un échantillon de taille n d!une loi μ de fonction de répartition F et $X_{(1)}, \ldots, X_{(n)}$ l'échantillon ordonné associé. Alors la fonction de répartition de $X_{(k)}$ est donnée par:*

$$F_k(t) = \frac{n!}{(k-1)!(n-k)!} \int_0^{F(t)} \theta^{k-1} (1-\theta)^{n-k} \, d\theta. \tag{4.31}$$

En particulier (formule facile à obtenir directement)

$$F_1(t) = 1 - (1 - F(t))^n, \quad F_n(t) = (F(t))^n. \tag{4.32}$$

Le cas le plus important est celui où μ a une densité p et dans ce cas:

Corollaire 4.9.2. *Soit $X_1, \ldots, X_n$ un n échantillon d'une loi sur $\mathbb{R}$ de densité $p(x)$ et de fonction de répartition F. Alors la densité de $X_{(k)}$ est donnée par:*

$$q_k(t) = \frac{n!}{(k-1)!(n-k)!} (F(t))^{k-1} (1 - F(t))^{n-k} p(t). \tag{4.33}$$

4.9.3. En fait lorsque μ a une densité p, il est facile de calculer la densité de l'échantillon ordonné en tant que loi marginale.

Théorème 4.9.3. *Soit $X_1, \ldots, X_n$ un n échantillon d'une loi sur $\mathbb{R}$ de densité $p(x)$. Alors la densité de $(X_{(1)}, \ldots, X_{(n)})$ est donnée par:*

$$f(x_1, \ldots, x_n) = n! \, p(x_1) \ldots p(x_n) 1_{x_1 < \ldots < x_n}. \tag{4.34}$$

Preuve: Soit $\mathfrak{S}_n$ l'ensemble des permutations de $\{1, 2, \ldots, n\}$. On a, pour $h \geq 0$,

$$\begin{aligned}
\mathbb{E}(h(X_{(1)}, \ldots, X_{(n)})) &= \sum_{\sigma \in \mathfrak{S}} \mathbb{E}(h(X_{(1)}, \ldots, X_{(n)}) 1_{\{X_{\sigma(1)} < \ldots < X_{\sigma(n)}\}}) \\
&= \sum_{\sigma \in \mathfrak{S}} \int_{\{x_{\sigma(1)} < \ldots < x_{\sigma(n)}\}} h(x_{\sigma(1)}, \ldots, x_{\sigma(n)}) p(x_1) \ldots p(x_n) \, dx_1 \ldots dx_n \\
&= \sum_{\sigma \in \mathfrak{S}} \int_{\{x_1 < \ldots < x_n\}} h(x_1, \ldots, x_n) p(x_1) \ldots p(x_n) \, dx_1 \ldots dx_n \\
&= n! \int_{\{x_1 < \ldots < x_n\}} h(x_1, \ldots, x_n) p(x_1) \ldots p(x_n) \, dx_1 \ldots dx_n. \ \diamond
\end{aligned}$$

A partir de (4.34), il est facile de retrouver (4.33) i.e. la densité q_k de $X_{(k)}$ considérée comme une marginale de $(X_{(1)}, \ldots, X_{(n)})$. On a donc, posant $A_t = \{x_1 < \ldots < x_{k-1} < t < x_{k+1} < \ldots < x_n\}$,

$$\begin{aligned}
q_k(t) &= n! \int_{A_t} p(x_1) \ldots p(x_{k-1}) p(t) p(x_{k+1}) \ldots p(x_n) \, dx_1 \ldots dx_{k-1} dx_{k+1} \ldots dx_n \\
&= \frac{n!}{(n-k)!} \int_{\{x_1 < \ldots < x_{k-1} < t\}} p(x_1) \ldots p(x_{k-1}) \, dx_1 \ldots dx_{k-1} (1 - F(t))^{n-k} p(t) \\
&= \frac{n!}{(k-1)!(n-k)!} \, (F(t))^{k-1} (1 - F(t))^{n-k} p(t).
\end{aligned}$$

Exemple. Soit $X_1, \ldots, X_n$ un n échantillon de la loi uniforme sur $[0, 1]$. Alors la loi de $(X_{(1)}, \ldots, X_{(n)})$ a pour densité $n! \, 1_{\{x_1 < \ldots < x_n\}}$ et celle de $X_{(k)}$, $1 \leq k \leq n$, a pour densité $\frac{n!}{(k-1)!(n-k)!} \, t^{k-1} (1-t)^{n-k} 1_{]0,1[}(t)$. En particulier (calcul facile en utilisant la formule (4.23)) $\mathbb{E}(X_{(k)}) = \frac{k}{n+1}$.

Chapitre 5

Fonctions caractéristiques. Vecteurs gaussiens

5.1. Transformée de Fourier

5.1.1. Rappelons que le produit de convolution de deux mesures bornées sur $\mathbb{R}^d$ a été défini en 3.5.4. Soient X et Y deux v.a. indépendantes à valeurs $\mathbb{R}^d$. On pose $S = X + Y$. Cherchons la loi de S. On a, pour toute $f \in \mathcal{B}^+(\mathbb{R}^d)$,

$$\mathbb{E}(f(S)) = \mathbb{E}(f(X+Y)) = \int f(x+y)\, d\mu_X(x) d\mu_Y(y) = \int f\, d\mu_X * \mu_Y.$$

On peut énoncer:

Proposition 5.1.1. *Soient X et Y deux v.a. indépendantes à valeurs $\mathbb{R}^d$. On a $\mu_{X+Y} = \mu_X * \mu_Y$.*

On sait que pour calculer des produits de convolution, la transformation de Fourier est un outil indispensable.

5.1.2. Transformée de Fourier. On note $\mathcal{M}_b$ l'ensemble des mesures bornées sur $\mathcal{B}(\mathbb{R}^d)$. Pour $\mu \in \mathcal{M}_b$, on pose

$$\hat{\mu}(t) = \int \mathrm{e}^{i<t,x>}\, d\mu(x), \quad t \in \mathbb{R}^d. \tag{5.1}$$

De même, pour $h \in L^1(\mathbb{R}^d, \lambda)$, λ mesure de Lebesgue sur $\mathbb{R}^d$, on pose

$$\hat{h}(t) = \int \mathrm{e}^{i<t,x>} h(x)\, dx, \quad t \in \mathbb{R}^d. \tag{5.2}$$

La fonction $\hat{\mu}$ (resp $\hat{h}$) s'appelle la transformée de Fourier de μ (resp. de h). Remarquer que, si $\mu = h.\lambda$, $\hat{\mu} = \hat{h}$. Alors,

Théorème 5.1.2. *(i) Soient $\mu,\nu \in \mathcal{M}_b$. Si $\hat{\mu} = \hat{\nu}$, $\mu = \nu$.*
(ii) Soit $\mu \in \mathcal{M}_b$ telle que $\hat{\mu} \in L^1(\lambda)$. On a alors $\mu = h.\lambda$ avec

$$h(x) = (2\pi)^{-d} \int e^{-i<t,x>} \hat{\mu}(t)\, dt. \tag{5.3}$$

Preuve: On pose:

$$g_\sigma(x) = (2\pi\sigma^2)^{-d/2} \exp(-\frac{|x|^2}{2\sigma^2}),\ \ |x|^2 = x_1^2 + \ldots + x_d^2. \tag{5.4}$$

Lemme 5.1.3. *La famille $(g_\sigma(x-a),\ \sigma > 0, a \in \mathbb{R}^d)$ est totale dans $C_0(\mathbb{R}^d)$.*

Preuve: Soit V l'espace vectoriel engendré par les fonctions $g_\sigma(x-a),\ \sigma > 0, a \in \mathbb{R}^d$. Vu que

$$g_\sigma(x-a)\, g_\rho(x-b) = C\, g_\tau(x-c) \text{ avec } \tau^2 = \frac{\rho^2\sigma^2}{\rho^2+\sigma^2},\ c = \frac{\rho^2 a + \sigma^2 b}{\rho^2+\sigma^2},$$

V est une algèbre. On vérifie immédiatement (i) et (ii) du th. 3.5.5 d'où $\overline{V} = C_0$. ◇

Lemme 5.1.4. *On a $\hat{g}_\sigma(t) = \exp(-\frac{\sigma^2}{2}|t|^2) = (2\pi\sigma^2)^{d/2} g_\sigma(\sigma^2 t)$.*

Preuve: Soit $\phi(t) = (2\pi)^{-1/2}\int \mathrm{e}^{itu}\mathrm{e}^{-u^2/2}\, du$, $t \in \mathbb{R}$. Vu que $|\frac{d}{dt}\mathrm{e}^{itu}| \le |u| \in L^1(\mathrm{e}^{-u^2/2}.\lambda)$, on peut appliquer la prop. 3.3.7 et on a

$$\phi'(t) = i(2\pi)^{-1/2}\int \mathrm{e}^{itu}\, d(-\mathrm{e}^{-u^2/2}) = -(2\pi)^{-1/2} t \int \mathrm{e}^{itu}\mathrm{e}^{-u^2/2}\, du = -t\phi(t)$$

d'où $\phi(t) = C\mathrm{e}^{-t^2/2} = \mathrm{e}^{-t^2/2}$ puisque $\phi(0) = 1$. Alors (th. 3.5.2)

$$(2\pi\sigma^2)^{-d/2}\int \mathrm{e}^{i<t,x>}\mathrm{e}^{-|x|^2/2\sigma^2}\, dx = \prod_{k=1}^{d} (2\pi\sigma^2)^{-1/2}\int \mathrm{e}^{it_k x_k}\mathrm{e}^{-x_k^2/2\sigma^2}\, dx_k = \mathrm{e}^{-\sigma^2|t|^2/2}.\ \diamond$$

Lemme 5.1.5. *Soit $\mu \in \mathcal{M}_b$. On a*

$$\int g_\sigma(x-a)\, d\mu(x) = (2\pi)^{-d/2}\int g_1(\sigma t)\, e^{-i<a,t>}\hat{\mu}(t)\, dt. \tag{5.5}$$

Si, de plus, $\hat{\mu} \in L^1(\lambda)$,

$$\int g_\sigma(x-a)\, d\mu(x) = (2\pi)^{-d}\int g_\sigma(x-a)\int e^{-i<x,t>}\hat{\mu}(t)\, dt\, dx. \tag{5.6}$$

Preuve: Notons d'abord que, vu le lem. 5.1.4,

$$g_\sigma(x) = (2\pi\sigma^2)^{-d/2}\hat{g}_\sigma(\frac{x}{\sigma^2}) = (2\pi)^{-d/2}\sigma^d \int g_\sigma(\sigma^2 t)\mathrm{e}^{i<x,t>}\, dt. \tag{5.7}$$

(i) On a, puisque $\int\int g_\sigma(\sigma^2 t)\,dtd\mu(x)<+\infty$,

$$\int g_\sigma(x-a)\,d\mu(x)=(2\pi)^{-d/2}\sigma^d\int\int g_\sigma(\sigma^2 t)\mathrm{e}^{i<x-a,t>}\,dtd\mu(x)$$

$$=(2\pi)^{-d/2}\sigma^d\int g_\sigma(\sigma^2 t)\mathrm{e}^{-i<a,t>}\int \mathrm{e}^{i<x,t>}d\mu(x)\,dt=(2\pi)^{-d/2}\sigma^d\int g_\sigma(\sigma^2 t)\mathrm{e}^{-i<a,t>}\hat{\mu}(t)\,dt$$

d'où (5.5) puisque $\sigma^d g_\sigma(\sigma^2 t)=g_1(\sigma t)$.

(ii) Si $\hat{\mu}\in L^1(\lambda)$, $g_\sigma(\sigma^2 u)\hat{\mu}(t)\in L^1(\lambda\otimes\lambda)$ et on a, vu que $g_\sigma(\sigma^2 t)=(2\pi\sigma^2)^{-d/2}\hat{g}_\sigma(t)$,

$$\int g_\sigma(x-a)\,d\mu(x)=(2\pi)^{-d/2}\sigma^d\int g_\sigma(\sigma^2 t)\mathrm{e}^{-i<a,t>}\hat{\mu}(t)\,dt$$

$$=(2\pi)^{-d}\int \mathrm{e}^{-i<a,t>}\hat{\mu}(t)\int \mathrm{e}^{i<u,t>}g_\sigma(u)\,dudt$$

$$=(2\pi)^{-d}\int g_\sigma(u)\int \mathrm{e}^{i<u-a,t>}\hat{\mu}(t)\,dtdu=(2\pi)^{-d}\int g_\sigma(x-a)\int \mathrm{e}^{-i<x,t>}\hat{\mu}(t)\,dtdx.$$

(On a posé $u=a-x$ et utilisé que $g_\sigma(-x)=g_\sigma(x)$). $\diamond$

Fin de la preuve. Soit $H=\{g_\sigma(x-a),\ \sigma>0, a\in\mathbb{R}^d\}$. Si $\hat{\mu}=\hat{\nu}$, on a, vu (5.5), pour toute $f\in H$, $\int f\,d\mu=\int f\,d\nu$ d'où, H étant total, $\mu=\nu$ (prop. 3.5.4 (iii)). De même, si $\hat{\mu}\in L^1$, posant $h(x)=(2\pi)^{-d}\int \mathrm{e}^{-i<x,t>}\hat{\mu}(t)\,dt$, on a vu (5.6), pour toute $f\in H$, $\int f\,d\mu=\int fh\,d\lambda$ d'où $\mu=h.\lambda$. $\diamond$

5.2. Fonctions caractéristiques

5.2.1. Soit X une v.a. à valeurs $\mathbb{R}^d$ de loi μ_X. On a alors, vu le th. 4.2.4, $\hat{\mu}_X(t)=\int \mathrm{e}^{i<t,x>}\,d\mu_X(x)=\mathbb{E}(\mathrm{e}^{i<t,X>})$. Ceci conduit à:

Définition 5.2.1. *Soit X une v.a. à valeurs $\mathbb{R}^d$. La fonction*

$$\phi_X(t)=\mathbb{E}(e^{i<t,X>})=\hat{\mu}_X(t)$$

s'appelle la fonction caractéristique de X.

Premières propriétés.

(i) ϕ_X est continue. En effet, si $t_n\to t$, $\mathrm{e}^{i<t_n,X>}\to\mathrm{e}^{i<t,X>}$ en ayant un module borné par 1. Il suffit d'appliquer le théorème de Lebesgue.

(ii) Pour $\alpha\in R$ et $b\in\mathbb{R}^d$, $\phi_{\alpha X+b}(t)=\mathrm{e}^{i<t,X>}\phi_X(\alpha t)$. En effet

$$\phi_{\alpha X+b}(t)=\mathbb{E}(\mathrm{e}^{i<t,\alpha X+b>})=\mathrm{e}^{i<t,b>}\mathbb{E}(\mathrm{e}^{i<t,\alpha X>})=\mathrm{e}^{i<t,b>}\mathbb{E}(\mathrm{e}^{i<\alpha t,X>}).$$

(iii) $\phi_{-X}(t)=\mathbb{E}(\mathrm{e}^{-i<t,X>})=\overline{\phi_X(t)}$.

(iv) Si $\mu_{-X}=\mu_X$ i.e. si la loi de X est symétrique, ϕ_X est réelle. Ceci résulte de (iii).

Le th. 5.1.2 devient:

Théorème 5.2.2. *Soient X et Y des v.a. à valeurs $\mathbb{R}^d$.*
(i) Si, pour tout t, $\phi_X(t) = \phi_Y(t)$, X et Y ont même loi.
(ii) Si $\phi_X \in L^1$, $\mu_X = h.\lambda$ avec

$$h(x) = (2\pi)^{-d} \int e^{-i<t,x>} \phi_X(t)\, dt.$$

Quant à prop. 5.1.1, elle s'énonce:

Théorème 5.2.3. *Soient X et Y deux v.a. indépendantes à valeurs $\mathbb{R}^d$. On a $\phi_{X+Y} = \phi_X \phi_Y$.*

Preuve: En fait cela se montre immédiatement grâce au th. 4.4.4:

$$\phi_{X+Y}(t) = \mathbb{E}(\mathrm{e}^{i<t,X+Y>}) = \mathbb{E}(\mathrm{e}^{i<t,X>}\mathrm{e}^{i<t,Y>}) = \mathbb{E}(\mathrm{e}^{i<t,X>})\mathbb{E}(\mathrm{e}^{i<t,Y>}) = \phi_X(t)\phi_Y(t). \diamond$$

5.2.2. Critère d'indépendance.

Théorème 5.2.4. *Des v.a. $X_1,\dots,X_n$ à valeurs $\mathbb{R}^{d_1},\dots,\mathbb{R}^{d_n}$ sont indépendantes ssi, pour tous $t_1 \in \mathbb{R}^{d_1},\dots,t_n \in \mathbb{R}^{d_n}$,*

$$\phi_{(X_1,\dots,X_n)}(t_1,\dots,t_n) = \phi_{X_1}(t_1)\dots\phi_{X_n}(t_n).$$

Preuve: En effet cette condition signifie que $\mu_{(X_1,\dots,X_n)}$ et $\mu_{X_1} \otimes \dots \otimes \mu_{X_n}$ ont même transformée de Fourier i.e. (th. 5.1.2) que $\mu_{(X_1,\dots,X_n)} = \mu_{X_1} \otimes \dots \otimes \mu_{X_n}$ ce qui équivaut (prop. 4.4.2) à l'indépendance de $X_1,\dots,X_n$. $\diamond$

5.2.3. Calcul des moments.

Proposition 5.2.5. *Soit X une v.a. à valeurs $\mathbb{R}^d$.*
(i) Si $X \in L^1_d$, ϕ_X est dérivable et $\frac{\partial \phi_X}{\partial t_k}(t) = \mathbb{E}(iX_k e^{i<t,X>})$. En particulier $\frac{\partial \phi_X}{\partial t_k}(0) = i\mathbb{E}(X_k)$.
(ii) Si $X \in L^2_d$, ϕ_X est deux fois dérivable et $\frac{\partial^2 \phi_X}{\partial t_j \partial t_k}(t) = -\mathbb{E}(X_j X_k e^{i<t,X>})$. En particulier $\frac{\partial^2 \phi_X}{\partial t_j \partial t_k}(0) = -\mathbb{E}(X_j X_k)$.

Preuve: (i) On remarque que $|\frac{\partial}{\partial t_k}\mathrm{e}^{i<t,X>}| = |X_k| \in L^1$ et on applique la prop. 3.3.7.

(ii) On continue.... $\diamond$

Il est facile de voir en appliquant la prop. 3.3.7 que si $X \in L^m_d$, ϕ_X est m fois dérivable et qu'on obtient les dérivées successives en dérivant sous le signe $\mathbb{E}$. Réciproquement on a ,

Proposition 5.2.6. *Soit X une v.a. à valeurs $\mathbb{R}^d$. Si ϕ_X est $2m$ fois dérivable en 0, m entier, $X \in L^{2m}_d$.*

Preuve: On se limite à $d = 1$, $m = 1$. On pose $\phi = \phi_X$ et $\mu = \mu_X$. On a $\phi"(0) = \lim_{h\to 0} \frac{1}{h^2}(\phi(h) + \phi(-h) - 2\phi(0))$ et

$$\phi(h) + \phi(-h) - 2\phi(0) = \int (\mathrm{e}^{ihx} + \mathrm{e}^{-ihx} - 2)\, d\mu(x) = -4\int \sin^2 \frac{hx}{2}\, d\mu(x).$$

Appliquant le lemme de Fatou (prop. 3.3.4), on a

$$-\phi"(0) = \lim_h 4\int \frac{\sin^2 \frac{hx}{2}}{h^2}\, d\mu(x) \geq 4\int \liminf_h \frac{\sin^2 \frac{hx}{2}}{h^2x^2} x^2\, d\mu(x) = \int x^2\, d\mu(x). \diamond$$

5.2.4. Fonctions caractéristiques usuelles (voir 2.2.5 et 4.3.1 pour les définitions).

a. Loi binomiale $B(n,p)$. Si $X \sim B(n,p)$, on a

$$\phi_X(t) = \mathbb{E}(\mathrm{e}^{itX}) = \sum_{k=0}^{n} C_n^k\, p^k(1-p)^{n-k}\mathrm{e}^{itk} = (p\mathrm{e}^{it} + 1 - p)^n.$$

Cette formule et le th. 5.2.3 montrent que, si $X \sim B(n,p)$ et $Y \sim B(m,p)$, X, Y indépendantes, alors $X + Y \sim B(n+m,p)$. En particulier si $X_1, \ldots, X_n$ sont des v.a. indépendantes avec $\mathbb{P}(X_k = 1) = p$, $\mathbb{P}(X_k = 0) = 1-p$, $S_n = X_1+\ldots+X_n \sim B(n,p)$.

b. Loi de Poisson $\mathcal{P}(\lambda)$. Si $X \sim \mathcal{P}(\lambda)$,

$$\phi_X(t) = \mathbb{E}(\mathrm{e}^{itX}) = \sum_{k=0}^{\infty} \mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}\mathrm{e}^{itk} = \exp(\lambda(\mathrm{e}^{it} - 1)).$$

Donc si $X \sim \mathcal{P}(\lambda)$ et $Y \sim \mathcal{P}(\mu)$, X, Y indépendantes, $X + Y \sim \mathcal{P}(\lambda + \mu)$.

c. Loi uniforme Si $X \sim U(a,b)$, $a < b$,

$$\phi_X(t) = \frac{1}{b-a}\int_{-a}^{b} \mathrm{e}^{itx}\, dx = \frac{\mathrm{e}^{itb} - \mathrm{e}^{ita}}{it(b-a)}.$$

d. Loi gamma $G(a,c)$. Si $X \sim G(a,c)$, on a

$$\phi_X(t) = \frac{c^a}{\Gamma(a)}\int_0^{+\infty} \mathrm{e}^{itx}\mathrm{e}^{-cx}x^{a-1}\, dx.$$

Utilisant la prop. 3.3.7 et intégrant par partie, on obtient

$$\phi'_X(t) = \frac{ic^a}{\Gamma(a)}\int_0^{+\infty} \mathrm{e}^{itx}\mathrm{e}^{-cx}x^{a}\, dx = \frac{-iac^a}{\Gamma(a)(it-c)}\int_0^{+\infty} \mathrm{e}^{itx}\mathrm{e}^{-cx}x^{a-1}\, dx = \frac{ia}{c-it}\phi_X(t)$$

d'où $\phi_X(t) = (1 - \frac{it}{c})^{-a}$ puisque $\phi_X(0) = 1$. Noter que pour $a \notin \mathbb{N}$, on prend la détermination continue valant 1 en 0.

Si $X \sim G(a,c)$ et $Y \sim G(b,c)$, X, Y indépendantes, alors $X+Y \sim G(a+b,c)$. En particulier si $X_1, \ldots, X_n$ sont des v.a. indépendantes de même densité $\lambda\mathrm{e}^{-\lambda x}1_{\mathbb{R}^+}$ et donc de loi $G(1,\lambda)$, $S_n = X_1+\ldots+X_n \sim G(n,\lambda)$ et a pour densité $\frac{\lambda^n}{(n-1)!}\mathrm{e}^{-\lambda x}x^{n-1}1_{\mathbb{R}^+}$.

e. Loi normale $N_1(m,\sigma^2)$. Si $Y \sim N_1(0,1)$, $\phi_Y(t) = \mathrm{e}^{-t^2/2}$ (lem. 5.1.4). Soit $X = m+\sigma Y$, alors $X \sim N_1(m,\sigma^2)$ et $\mathbb{E}(\mathrm{e}^{itX}) = \mathrm{e}^{mt}E(\mathrm{e}^{t\sigma Y})$, d'où la formule:

$$\phi_X(t) = \exp(itm - \frac{1}{2}\sigma^2 t^2),\ X \sim N_1(m,\sigma^2). \tag{5.8}$$

On en déduit immédiatement

Proposition 5.2.7. *Si $X \sim N_1(m,\sigma^2)$ et $Y \sim N_1(l,\rho^2)$, X, Y indépendantes, alors $X+Y \sim N_1(m+l, \sigma^2+\rho^2)$.*

f. Loi de Laplace. C'est la loi d'une v.a. X de densité $q(x) = \frac{1}{2}\mathrm{e}^{-|x|}$. On a

$$\phi_X(t) = \frac{1}{2}\int_{-\infty}^{+\infty} \mathrm{e}^{itx}\mathrm{e}^{-|x|}\,dx = \frac{1}{2}\int_0^{+\infty} \mathrm{e}^{x(it-1)}\,dx + \frac{1}{2}\int_{-\infty}^{0} \mathrm{e}^{x(it+1)}\,dx = \frac{1}{1+t^2}.$$

g. Loi de Cauchy de paramètre 0. C'est la loi d'une v.a. X de densité $q(x) = \frac{1}{\pi(1+x^2)}$. Vu que $\frac{1}{1+t^2} \in L^1$, on a d'après f. et le th. 5.2.2 (ii),

$$\frac{1}{2\pi}\int_{-\infty}^{+\infty} \mathrm{e}^{-itx}\frac{1}{1+t^2}\,dt = \frac{1}{2}\mathrm{e}^{-|x|}.$$

On en déduit

$$\phi_X(t) = \frac{1}{\pi}\int_{-\infty}^{+\infty} \mathrm{e}^{itx}\frac{1}{1+x^2}\,dt = \mathrm{e}^{-|t|}.$$

5.3. Vecteurs gaussiens

5.3.1. On dit qu'une probabilité μ sur $\mathbb{R}$ est gaussienne si elle a pour densité (4.13) ou si $\mu = \delta_m$. Il est normal d'adjoindre les mesures de Dirac aux lois gaussiennes car la loi $N_1(m,\sigma^2)$ converge en un certain sens vers δ_m lorsque $\sigma \to 0$. Une v.a. réelle est dite gaussienne si sa loi est gaussienne.

Définition 5.3.1. *Un vecteur aléatoire $X = (X_1,\ldots,X_d)$ est dit gaussien si, pour tout $a \in \mathbb{R}^d$, $a^{\mathrm{T}}X = a_1X_1+\ldots+a_dX_d$ est une v.a. gaussienne.*

En particulier chaque composante X_k est une v.a.r. gaussienne mais cela ne suffit pas à assurer que le vecteur X est gaussien.

On appelle loi gaussienne sur $\mathbb{R}^d$ toute loi d'un vecteur gaussien.

Exemples. (i) $X = 0 \in \mathbb{R}^d$ est un vecteur gaussien.

(ii) Soit $X = (X_1,\ldots,X_d)$ avec $X_1,\ldots,X_d$ indépendants de même loi $N_1(0,1)$. Alors (prop. 5.2.7) $a_1X_1+\ldots+a_dX_d \sim N_1(0, a_1^2+\ldots+a_d^2)$ et X est un vecteur gaussien.

Cette notion est invariante par transformation linéaire, plus précisément:

Lemme 5.3.2. *Soit X un vecteur gaussien à valeurs $\mathbb{R}^d$ de moyenne m et de matrice de covariance D. Pour tous $b \in \mathbb{R}^r$ et M matrice $r \times d$, $Y = b+MX$ est un vecteur gaussien à valeurs $\mathbb{R}^r$, de moyenne $b+Mm$ et de matrice de covariance MDM^{T}*

Preuve: En effet $a^{\mathsf T}Y = a^{\mathsf T}b + (a^{\mathsf T}M)X$ est une v.a.r. gaussienne. On a $\mathbb{E}(Y) = b + M\mathbb{E}(X) = b + Mm$ et (prop. 4.5.4) $K(Y) = K(MX) = MK(X)M^{\mathsf T} = MDM^{\mathsf T}$. ⋄

Théorème 5.3.3. *Soit X un vecteur aléatoire de moyenne m et de matrice de covariance K. Le vecteur X est gaussien ssi sa fonction caractéristique est donnée par*

$$\phi_X(t) = \exp(it^{\mathsf T}m - \frac{1}{2}t^{\mathsf T}Kt). \tag{5.9}$$

Preuve: (i) Supposons X gaussien. Alors (lem. 5.3.2) $t^{\mathsf T}X \sim N_1(t^{\mathsf T}m, t^{\mathsf T}Kt)$ et $\phi_X(t) = \mathbb{E}(e^{it^{\mathsf T}X}) = \phi_{t^{\mathsf T}X}(1) = \exp(it^{\mathsf T}m - \frac{1}{2}t^{\mathsf T}Kt)$ d'où (5.9).

(ii) Supposons (5.9). Alors $\phi_{a^{\mathsf T}X}(u) = \mathbb{E}(e^{iua^{\mathsf T}X}) = \exp(iua^{\mathsf T}m - \frac{1}{2}u^2a^{\mathsf T}Ka)$ donc $a^{\mathsf T}X$ est une v.a.r. gaussienne et X un vecteur gaussien. ⋄

Toute loi gaussienne sur $\mathbb{R}^d$ est donc déterminée par sa moyenne m et sa matrice de covariance K. On note $N_d(m, K)$ une telle loi. On a vu (exemple (ii)) que $N_d(0, I_d)$ existe mais on n'a pas établi l'existence dans le cas général. Pour cela, on utilise:

Lemme 5.3.4. *Soit K une matrice $d \times d$ symétrique semi-définie positive. Il existe une matrice $d \times d$ symétrique semi-définie positive A telle que $K = A^2$.*

Preuve: Soient $\lambda_1, \ldots, \lambda_d$ les valeurs propres de K (elles sont ≥ 0). Il existe une matrice orthogonale C (i.e. $C^{\mathsf T}C = I$) telle que $C^{\mathsf T}KC = D = \mathrm{diag}(\lambda_1, \ldots, \lambda_d)$ où $\mathrm{diag}(\lambda_1, \ldots, \lambda_d)$ désigne la matrice diagonale ayant $\lambda_1, \ldots, \lambda_d$ sur la diagonale. On a alors $CDC^{\mathsf T} = K$. Soit $\Delta = \mathrm{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_d})$. On pose $A = C\Delta C^{\mathsf T}$. On a,

$$A^2 = C\Delta C^{\mathsf T}C\Delta C^{\mathsf T} = C\Delta^2 C^{\mathsf T} = CDC^{\mathsf T} = K. \diamond$$

Appliquant le lem. 5.3.2, on a que, si $X \sim N_d(0, I_d)$, $Y = m + AX \sim N_d(m, K)$. On a montré:

Théorème 5.3.5. *Etant donnés $m \in \mathbb{R}^d$ et une matrice $d \times d$ symétrique semi-définie positive K, il existe une et une seule loi gaussienne sur $\mathbb{R}^d$ de moyenne m et de matrice de covariance K.*

5.3.2. Vecteurs gaussiens et indépendance.

Théorème 5.3.6. *Soient $X = (X_1, \ldots, X_d)$ un vecteur gaussien.*
(i) Les v.a.r. $X_1, \ldots, X_d$ sont indépendantes ssi la matrice de covariance $K(X)$ est diagonale.
(ii) On pose

$$Y_1 = (X_1, \ldots, X_{d_1}),\ Y_2 = (X_{d_1+1}, \ldots, X_{d_2}), \ldots Y_r = (X_{d_{r-1}+1}, \ldots, X_d).$$

Les vecteurs $(Y_1, \ldots, Y_r)$ sont indépendants ssi $K_{ij}(X) = Cov(X_i, X_j) = 0$ pour tous i, j n'appartenant pas au même intervalle $[1, d_1]$, $[d_1 + 1, d_2], \ldots, [d_{r-1} + 1, d]$.

Preuve: Seule la suffisance demande une preuve.

(i) Supposons $K(X)$ diagonale. On a $K(X) = \text{diag}(\sigma_1^2, \dots, \sigma_d^2)$ où $\sigma_k^2 = \text{Var}(X_k)$. Alors, notant $m = \mathbb{E}(X)$,

$$\phi_X(t) = \exp(i\sum_{k=1}^d m_k t_k - \frac{1}{2}\sum_{k=1}^d \sigma_k^2 t_k^2) = \prod_{k=1}^d \exp(i m_k t_k - \frac{1}{2}\sigma_k^2 t_k^2) = \phi_{X_1}(t_1)\dots\phi_{X_d}(t_d)$$

et donc (prop. 5.2.4) les X_k sont indépendantes.

(ii) Supposons la condition sur les covariances réalisées. Elle implique, pour tous $u_1 \in \mathbb{R}^{d_1}, u_2 \in \mathbb{R}^{d_2-d_1}, \dots$ et $p \neq q$, $\text{Cov}(u_p^{\mathsf{T}} Y_p, u_q^{\mathsf{T}} Y_q) = 0$. Donc, d'après (i), les v.a.r. $u_1^{\mathsf{T}} Y_1, \dots, u_r^{\mathsf{T}} Y_r$ sont indépendantes. On a alors

$$\mathbb{E}(\mathrm{e}^{i(u_1^{\mathsf{T}} Y_1 + \dots + u_r^{\mathsf{T}} Y_r)}) = \mathbb{E}(\mathrm{e}^{i u_1^{\mathsf{T}} Y_1}) \dots \mathbb{E}(\mathrm{e}^{i u_r^{\mathsf{T}} Y_r})$$

et (prop. 5.2.4) les v.a. $Y_1, \dots, Y_r$ sont indépendantes. ⋄

Remarque. Attention à l'utilisation du th. 5.3.6. On peut avoir X et Y v.a.r. gaussiennes, $\text{Cov}(X,Y) = 0$ sans que les v.a. X et Y soient indépendantes. Par exemple si $X \sim N_1(0,1)$, si U est une v.a. indépendante de X telle que $\mathbb{P}(U = 1) = \mathbb{P}(U = -1) = \frac{1}{2}$ et si $Y = UX$, on vérifie facilement que $Y \sim N_1(0,1)$. On a $\text{Cov}(X,Y) = \mathbb{E}(XY) = \mathbb{E}(UX^2) = \mathbb{E}(U)\mathbb{E}(X^2) = 0$ et $|X| = |Y|$ donc X et Y ne sont pas indépendantes. En fait le couple (X,Y) n'est pas gaussien.

5.3.3. Le cas non dégénéré. On dit que la loi $N_d(m,K)$ est non dégénérée si $\det(K) \neq 0$. Dans ce cas:

Théorème 5.3.7. *Si $X \sim N_d(m,K)$ et si $det(K) \neq 0$, X admet la densité*

$$h_{m,K}(x) = (2\pi)^{-\frac{d}{2}} (det(K))^{-\frac{1}{2}} \exp(-\frac{1}{2}(x-m)^{\mathsf{T}} K^{-1}(x-m)).$$

Preuve: Soit A une matrice $d \times d$ telle que $K = AA^{\mathsf{T}}$, on a $\det(A) = (\det(K))^{1/2}$ et A est inversible. Soit $Y \sim N_d(0, I_d)$ un vecteur gaussien de densité $(2\pi)^{-d/2}\exp(-\frac{|y|^2}{2})$. On a (lem . 5.3.2) $Y = m + AY \sim N_d(m,K)$ et, pour $f \in \mathcal{B}^+(\mathbb{R}^d)$,

$$\mathbb{E}(f(X)) = \mathbb{E}(f(m+AY)) = (2\pi)^{-\frac{d}{2}} \int f(m+Ay)\exp(-\frac{|y|^2}{2})\,dy.$$

On effectue le changement de variable $y = A^{-1}(x-m)$, on a $\frac{D(y)}{D(x)} = \det(A^{-1})$ et

$$\mathbb{E}(f(X)) = (2\pi)^{-\frac{d}{2}}\det(A^{-1}) \int f(x)\exp(-\frac{1}{2}(x-m)^{\mathsf{T}}(A^{-1})^{\mathsf{T}}A^{-1}(x-m))\,dx.$$

Comme $K^{-1} = (AA^{\mathsf{T}})^{-1} = (A^{-1})^{\mathsf{T}}A^{-1}$, on a la formule annoncée. ⋄

Chapitre 6

Convergence des suites de variables aléatoires

6.1. Modes de convergence

6.1.1. Principaux modes de convergence.

Définition 6.1.1. *Soient X_n et X des v.a. à valeurs $\mathbb{R}^d$.*
(i) On dit que X_n converge en probabilité vers X si, pour tout $\varepsilon > 0$, $\mathbb{P}(|X_n - X| > \varepsilon) \to_n 0$.
(ii) On dit que X_n converge presque sûrement (en abrégé p.s.) vers X si, pour tout $\omega \notin N$, N négligeable, $X_n(\omega) \to_n X(\omega)$.
(iii) On dit que X_n converge vers X dans L^p, $1 \le p < +\infty$, si X_n et X sont dans L^p et si $\mathbb{E}(|X_n - X|^p) \to_n 0$.

La convergence dans L^1 s'appelle aussi la convergence en moyenne, la convergence dans L^2 s'appelle aussi la convergence en moyenne quadratique. On vérifie immédiatement que $X_n = (X_n^1, \dots, X_n^d)$ converge vers $X = (X^1 \dots, X^d)$ en un des sens ci-dessus ssi, pour $k = 1, \dots, d$, X_n^k converge vers X^k dans le même sens. On ne considérera donc plus que des v.a. réelles.

Rappelons qu'on note, pour X v.a.r., $||X||_p = (\mathbb{E}|X|^p)^{\frac{1}{p}}$. Vu l'inégalité de Hölder (3.7), on a, pour $1 \le p \le q$, $||X||_p \le ||X||_q$ et donc la convergence dans L^q implique la convergence dans L^p. En particulier la convergence dans L^2 implique la convergence dans L^1.

Proposition 6.1.2. *La convergence dans L^1 implique la convergence en probabilité, la convergence p.s. implique la convergence en probabilité.*

Preuve: (i) D'après l'inégalité de Markov (prop. 4.2.7), $\mathbb{P}(|X_n - X| > \varepsilon) \le \varepsilon^{-1}\mathbb{E}(|X_n - X|)$ ce qui montre le premier point.

(ii) Supposons que X_n converge p.s. vers X. Alors, pour tout $\varepsilon > 0$, $1_{\{|X_n-X|>\varepsilon\}} \to_n 0$ p.s. et est manifestement borné par 1, donc (th. de Lebesgue) $\mathbb{P}(|X_n - X| > \varepsilon) = \mathbb{E}(1_{\{|X_n-X|>\varepsilon\}}) \to_n 0$. $\diamond$

Notons que si X_n converge en probabilité vers X et vers Y, on a $\mathbb{P}(|X - Y| > \varepsilon) \leq \mathbb{P}(|X - X_n| > \frac{\varepsilon}{2}) + \mathbb{P}(|X_n - Y| > \frac{\varepsilon}{2}) \to_n 0$ et donc $\mathbb{P}(|X - Y| > 0) = 0$ et $X = Y$ p.s. Ceci implique, vu la prop. 6.1.2, que les limites de X_n en les différents sens définis ci-dessus sont p.s. égales.

6.1.2. Exemples. Soit X_n une suite de v.a.r. indépendantes telles que $\mathbb{P}(X_n = a_n) = p_n$, $\mathbb{P}(X_n = 0) = 1 - p_n$. On suppose $0 < p_n < 1$, $p_n \to_n 0$ et $a_n \geq 1$.

a. On a, pour $\varepsilon \in]0, 1[$, $\mathbb{P}(|X_n| > \varepsilon) = \mathbb{P}(X_n > \varepsilon) = p_n$ et $X_n \to_n 0$ en probabilité.

b. On a $\sum \mathbb{P}(X_n > 0) = \sum p_n$ donc, si $\sum p_n < +\infty$, on a (prop. 4.1.2) que $\{X_n > 0\}$ n'a p.s. lieu que pour un nombre fini de n donc $X_n \to_n 0$ p.s. Réciproquement si $\sum p_n = +\infty$, on a (prop. 4.1.2) que $\{X_n = a_n\}$ a p.s. lieu pour une infinité de n donc X_n ne converge pas p.s. vers 0. Donc $X_n \to_n 0$ p.s. ssi $\sum p_n < +\infty$.

c. $\mathbb{E}|X_n| = \mathbb{E}(X_n) = a_n p_n$. Donc $X_n \to_n 0$ dans L^1 ssi $a_n p_n \to_n 0$.

d. $\mathbb{E}(X_n)^2 = a_n^2 p_n$. Donc $X_n \to_n 0$ dans L^2 ssi $a_n^2 p_n \to_n 0$.

Si on choisit $p_n = \frac{1}{n}$, $a_n = 1$, X_n converge vers 0 dans L^1 mais pas p.s. Si on choisit $p_n = \frac{1}{n^2}$, $a_n = n^2$, X_n converge vers 0 p.s. mais pas dans L^1. Si on choisit $p_n = \frac{1}{n^2}$, $a_n = n$, X_n converge vers 0 dans L^1 mais pas dans L^2.

6.1.3. Critères de convergence.

Proposition 6.1.3. *Soit X_n une suite de v.a.r. Si $\sum \mathbb{P}(|X_{n+1} - X_n| > \varepsilon_n) < +\infty$ pour une suite $\varepsilon_n > 0$ vérifiant $\sum \varepsilon_n < +\infty$, la suite X_n converge p.s.*

Preuve: D'après le lemme de Borel-Cantelli (prop. 4.1.2), pour tout $\omega \notin N$, N négligeable, il existe $n_0(\omega)$ tel que, pour tout $n \geq n_0(\omega)$, $|X_{n+1}(\omega) - X_n(\omega)| \leq \varepsilon_n$. On a donc, pour $n > m \geq n_0(\omega)$,

$$|X_n(\omega) - X_m(\omega)| \leq \sum_{k=m}^{n-1} |X_{k+1}(\omega) - X_k(\omega)| \leq \sum_{k=m}^{n-1} \varepsilon_k.$$

Vu la convergence de $\sum \varepsilon_n$, ceci implique que $X_n(\omega)$ est une suite de Cauchy et donc $X_n(\omega)$ converge. $\diamond$

Corollaire 6.1.4. *De toute suite X_n convergeant en probabilité, on peut extraire une sous-suite X_{n_k} convergeant p.s.*

Preuve: Vu que, pour tout k, $\mathbb{P}(|X_n - X| > 2^{-(k+1)}) \to_n 0$, on peut construire une suite croissante n_k telle que, pour tout $n \geq n_k$, $\mathbb{P}(|X_n - X| > 2^{-(k+1)}) \leq 2^{-(k+1)}$. On a alors,

$$\mathbb{P}(|X_{n_{k+1}} - X_{n_k}| > 2^{-k}) \leq \mathbb{P}(|X_{n_{k+1}} - X| > 2^{-(k+1)}) + \mathbb{P}(|X_{n_k} - X| > 2^{-(k+1)}) \leq 2^{-k}.$$

D'où (prop. 6.1.3) X_{n_k} converge p.s. $\diamond$

Il est très utile d'avoir des critères de type Cauchy.

Proposition 6.1.5. *Soit X_n une suite de v.a.r.*
(i) X_n converge en probabilité ssi, pour tout $\rho > 0$, $\sup_k \mathbb{P}(|X_{n+k} - X_n| > \rho) \to_n 0$,
(ii) X_n converge dans L^p $(1 \le p < +\infty)$ ssi $\sup_k \mathbb{E}(|X_{n+k} - X_n|^p) \to_n 0$,
(iii) X_n converge p.s. ssi, pour tout $\rho > 0$, $\mathbb{P}(\sup_k |X_{n+k} - X_n| > \rho) \to_n 0$.

Preuve: (i) Supposons que, pour tout $\rho > 0$, $\sup_k \mathbb{P}(|X_{n+k} - X_n| > \rho) \to_n 0$. On peut alors construire une suite croissante d'entiers n_r telle que $\mathbb{P}(|X_{n_r+1} - X_{n_r}| > 2^{-r}) \le 2^{-r}$ et donc (prop. 6.1.3) X_n converge p.s. et a fortiori en probabilité vers une v.a. X. Alors, étant donné $\varepsilon > 0$,

$$\mathbb{P}(|X_n - X| > \rho) \le \mathbb{P}(|X_n - X_{n_r}| > \rho/2) + \mathbb{P}(|X - X_{n_r}| > \rho/2) < \varepsilon$$

pour tout $n \ge n_r$ si on choisit r assez grand et $X_n \to X$ en probabilité. Vu que $\mathbb{P}(|X_{n+k} - X_n| > \rho) \le \mathbb{P}(|X_{n+k} - X| > \rho/2) + \mathbb{P}(|X_n - X| > \rho/2)$, la réciproque est immédiate.

(ii) Ceci n'est rien d'autre que la complétude de L^p (voir 3.3.5).

(iii) Supposons que, pour tout $\rho > 0$, $\mathbb{P}(\sup_k |X_{n+k} - X_n| > \rho) \to_n 0$. Soit $V_n = \sup_{i,j \ge n} |X_i - X_j|$, alors $V_n \downarrow V$ et X_n converge p.s. ssi $V = 0$ p.s. (critère de Cauchy). Mais $\mathbb{P}(V_n > \rho) \le \mathbb{P}(\sup_{k \ge 1} |X_{n+k} - X_n| > \rho/2) \to_n 0$ ce qui implique que $V = 0$ p.s. Réciproquement si X_n converge p.s., $\sup_k |X_{n+k} - X_n| \to_n 0$ p.s. et aussi en probabilité. $\diamond$

6.2. Loi 0-1

6.2.1. Soit $X_1, \ldots, X_n, \ldots$ une suite de v.a. à valeurs $\mathbb{R}^d$. On pose:

$$\mathcal{F}_n(X) = \sigma(X_1, \ldots, X_n), \ \ \mathcal{F}_\infty(X) = \sigma(X_1, \ldots, X_n, \ldots) = \sigma(\cup_{n \ge 1} \mathcal{F}_n(X)),$$
$$\mathcal{F}^n(X) = \sigma(X_n, X_{n+1}, \ldots, X_{n+k}, \ldots), \ \ \mathcal{F}^\infty(X) = \cap_{n \ge 1} \mathcal{F}^n(X).$$

Evidemment $\mathcal{F}^\infty(X) \subset \mathcal{F}_\infty(X)$. La tribu $\mathcal{F}^\infty(X)$ s'appelle la tribu asymptotique ou tribu de queue de la suite X_n.

Exemple. Soit $X_1, \ldots, X_n, \ldots$ une suite de v.a. réelles. Les événements

$$\{\sum X_n \text{ converge}\}, \ \ \{\sum |X_n| < +\infty\}, \ \ \{\limsup \frac{1}{n}(X_1 + \ldots + X_n) < 1\}$$

sont dans $\mathcal{F}^\infty(X)$. En effet il suffit de vérifier que, pour tout p, ils sont dans $\mathcal{F}^p$, ce qui est immédiat.

6.2.2. En fait, si les X_n sont indépendantes, un événement de $\mathcal{F}^\infty(X)$ est de probabilité 0 ou 1. C'est la loi 0-1.

Proposition 6.2.1. *Soit $X_1, \ldots, X_n, \ldots$ une suite de v.a. indépendantes à valeurs $\mathbb{R}^d$. Alors, pour tout $A \in \mathcal{F}^\infty(X)$, $\mathbb{P}(A) = 0$ ou 1. De plus, si Y est une v.a.r. $\mathcal{F}^\infty(X)$-mesurable, $Y =$ constante p.s.*

Preuve: Soit $A \in \mathcal{F}^\infty(X)$ avec $\mathbb{P}(A) > 0$. On pose

$$\mathbb{Q}(B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}, \quad B \in \mathcal{F}_\infty(X).$$

$\mathbb{Q}$ est une probabilité sur $\mathcal{F}_\infty(X)$. Si $B \in \mathcal{F}_n(X)$, B et A sont indépendants puisque $A \in \mathcal{F}^{n+1}(X)$. On a donc $\mathbb{P}(B) = \mathbb{Q}(B)$ pour tout $B \in \mathcal{C} = \cup_{n\geq 1}\mathcal{F}_n(X)$. Cette classe étant stable par intersection finie et engendrant $\mathcal{F}_\infty(X)$, on a (cor. 3.2.3) $\mathbb{P}(B) = \mathbb{Q}(B)$ pour tout $B \in \mathcal{F}_\infty(X)$ et en particulier pour $B = A$. Donc $\mathbb{P}(A) = \mathbb{Q}(A) = 1$.

Soit $F_Y(t) = \mathbb{P}(Y \leq t)$. Par hypothèse, $\{Y \leq t\} \in \mathcal{F}^\infty(X)$ et donc $F_Y(t) = 0$ ou 1 ce qui implique qu'il existe $a \in \mathbb{R}$ tel que $F_Y(t) = 1_{[a,+\infty[}(t)$ et donc $Y = a$ p.s. $\diamond$

Corollaire 6.2.2. *Soit $X_1, \ldots, X_n, \ldots$ une suite de v.a. réelles indépendantes. Alors,*
(i) $\sum X_n$ converge p.s. ou diverge p.s.,
(ii) si b_n est une suite de réels tendant vers $+\infty$, $\frac{1}{b_n}(X_1 + \ldots + X_n)$ diverge p.s. ou converge vers une constante p.s.

Preuve: On a vu que $\{\sum X_n \text{ converge}\} \in \mathcal{F}^\infty(X)$ d'où (i). De même $A = \{\frac{1}{b_n}(X_1 + \ldots + X_n) \text{ converge}\} \in \mathcal{F}^\infty(X)$ donc $\mathbb{P}(A) = 0$ ou 1. Supposons que $\mathbb{P}(A) = 1$. Soit $Z = \lim_n \frac{1}{b_n}(X_1 + \ldots + X_n)$. Vu que $b_n \to_n +\infty$, on a aussi, pour tout p, $Z = \lim_n \frac{1}{b_n}(X_p + \ldots + X_n)$ et donc $Z \in [\mathcal{F}^\infty(X)]$ d'où $Z =$ constante p.s. $\diamond$

6.3. Somme de v.a. indépendantes

Soit $X_1, \ldots, X_n, \ldots$ une suite de v.a. réelles de carré intégrable. On pose $S_n = X_1 + \ldots + X_n$ et $Y_n = X_n - \mathbb{E}(X_n)$. On a alors

$$S_n = \sum_{k=1}^{n} Y_k + \sum_{k=1}^{n} \mathbb{E}(X_k) \tag{6.1}$$

et $\mathbb{E}(Y_k) = 0$, $\mathbb{E}(Y_k^2) = \mathrm{Var}(Y_k) = \mathrm{Var}(X_k)$. Donc pour étudier la convergence de S_n, il suffit pour l'essentiel de s'intéresser au cas centré.

6.3.1. La convergence dans L^2 est simple à étudier.

Proposition 6.3.1. *Soit $X_1, \ldots, X_n, \ldots$ une suite de v.a. réelles, indépendantes, de carré intégrable et centrées. Alors S_n converge dans L^2 ssi la série $\sum \mathbb{E}(X_n^2)$ est convergente.*

Preuve: On a, pour $n < m$,

$$\mathbb{E}[(S_m - S_n)^2] = \mathbb{E}(\sum_{k=n+1}^{m} X_k)^2 = \sum_{k=n+1}^{m} \mathbb{E}(X_k^2).$$

On en déduit que S_n est une suite de Cauchy de L^2 et donc converge dans L^2 ssi $\sum \mathbb{E}(X_n^2) < +\infty$. ⋄

6.3.2. L'outil de base est l'inégalité suivante due à Kolmogorov.

Proposition 6.3.2. *Soit $X_1, \ldots, X_n, \ldots$ une suite de v.a. réelles, indépendantes, de carré intégrable et centrées. Alors, pour tout $\rho > 0$ et tout n,*

$$\mathbb{P}(\max_{1\leq k\leq n} |S_k| \geq \rho) \leq \frac{1}{\rho^2}\sum_{k=1}^{n} \mathbb{E}(X_k^2).$$

Preuve: On pose $A = \{\max_{1\leq k\leq n} |S_k| \geq \rho\}$ et, pour $k = 1, \ldots, n$, $B_k = \{|S_1| < \rho, \ldots, |S_{k-1}| < \rho, |S_k| \geq \rho\}$. Les ensembles B_k sont disjoints d'union A. Noter que, pour $k \leq n$,

$$\mathbb{E}(1_{B_k} S_n^2) = \mathbb{E}(1_{B_k}(S_k + S_n - S_k)^2) = \mathbb{E}(1_{B_k} S_k^2) + \mathbb{E}(1_{B_k}(S_n - S_k)^2) \geq \mathbb{E}(1_{B_k} S_k^2)$$

puisque, les v.a. $1_{B_k} S_k$ et $S_n - S_k$ étant indépendantes,

$$\mathbb{E}(1_{B_k} S_k(S_n - S_k)) = \mathbb{E}(1_{B_k} S_k)\mathbb{E}(S_n - S_k) = 0.$$

On a alors, vu que $S_k^2 \geq \rho^2$ sur B_k,

$$\rho^2\mathbb{P}(A) = \rho^2\sum_{k=1}^{n}\mathbb{P}(B_k) \leq \sum_{k=1}^{n}\mathbb{E}(1_{B_k} S_k^2) \leq \sum_{k=1}^{n}\mathbb{E}(1_{B_k} S_n^2) \leq \mathbb{E}(S_n^2) = \sum_{k=1}^{n}\mathbb{E}(X_k^2). \diamond$$

6.3.3. On peut maintenant établir le résultat principal.

Théorème 6.3.3. *Soit $X_1, \ldots, X_n, \ldots$ une suite de v.a. réelles, indépendantes, de carré intégrable. Si les séries $\sum \mathbb{E}(X_n)$ et $\sum$ Var(X_n) convergent, S_n converge p.s. et dans L^2.*

Preuve: Supposons d'abord les X_n centrées. Appliquant la prop. 6.3.2 à la suite $X_{m+1}, \ldots, X_{m+k}, \ldots$, on a

$$\mathbb{P}(\max_{1\leq k\leq n} |S_{m+k} - S_m| > \rho) = \mathbb{P}(\max_{1\leq k\leq n} |\sum_{i=1}^{k} X_{m+i}| > \rho) \leq \frac{1}{\rho^2}\sum_{i=1}^{n}\mathbb{E}(X_{m+i}^2) = \frac{1}{\rho^2}\sum_{k=m+1}^{m+n}\mathbb{E}(X_k^2).$$

On en déduit

$$\mathbb{P}(\sup_{k\geq 1} |S_{m+k} - S_m| > \rho) = \lim_{n\to+\infty}\mathbb{P}(\max_{1\leq k\leq n} |S_{m+k} - S_m| > \rho) \leq \frac{1}{\rho^2}\sum_{k>m}\mathbb{E}(X_k^2) \to_m 0.$$

Donc (prop. 6.1.5) S_n converge p.s. et aussi (prop. 6.3.1) dans L^2.

Pour le cas général, il suffit d'utiliser (6.1). ⋄

Remarque. On peut se demander si le th. 6.3.3 admet une réciproque. Sans hypothèse supplémentaire, il n'en est rien. En effet, soit $X_1, \ldots, X_n, \ldots$ une suite de

v.a.r. indépendantes telles que $\mathbb{P}(X_n = a_n) = p_n$, $\mathbb{P}(X_n = -a_n) = p_n$ et $\mathbb{P}(X_n = 0) = 1 - 2p_n$ avec $a_n > 0$, $0 < p_n < \frac{1}{2}$. On a $\sum_n \mathbb{P}(X_n \neq 0) = 2\sum_n p_n$. Donc si $\sum_n p_n < +\infty$, d'après Borel-Cantelli, p.s. $X_n = 0$ à partir d'un certain rang et $S_n = \sum_{k=1}^n X_k$ converge p.s. alors qu'on peut avoir $\sum \mathbb{E}(X_n^2) = 2\sum p_n a_n^2 = +\infty$ (prendre par exemple $p_n = n^{-2}$ et $a_n = n$). Pour plus de précisions, voir 6.5.

6.3.4. On s'intéresse maintenant à la convergence de $\frac{S_n}{b_n}$, b_n étant une suite tendant vers $+\infty$. On se ramène au cas précédent grâce au lemme de Kronecker:

Lemme 6.3.4. *Soient, pour $n \geq 1$, $b_n, x_n \in \mathbb{R}$, $0 < b_n \uparrow_n +\infty$ et $s_n = x_1 + \ldots + x_n$. Si la série $\sum \frac{x_n}{b_n}$ converge, $\frac{s_n}{b_n} \to 0$.*

Preuve: On pose $b_0 = 0$, $v_n = b_n - b_{n-1}$, $z_0 = 0$, $z_n = \sum_{k=1}^n \frac{x_k}{b_k}$. On a donc $b_n = \sum_{k=1}^n v_k$ et

$$\sum_{k=1}^n x_k = \sum_{k=1}^n b_k(z_k - z_{k-1}) = b_n z_n - \sum_{k=1}^n v_k z_{k-1} = \sum_{k=1}^n v_k(z_n - z_k).$$

On en déduit que, pour tout $p < n$,

$$|\frac{1}{b_n}\sum_{k=1}^n x_k| \leq \frac{1}{b_n}|\sum_{k=1}^p v_k(z_n - z_{k-1})| + \frac{1}{b_n}(\sum_{k=p+1}^n v_k) \max_{p\leq k\leq n} |z_n - z_{k-1}|.$$

D'où, puisque $b_n \to_n +\infty$ et $\frac{1}{b_n}(\sum_{k=p+1}^n v_k) \leq 1$, pour tout p,

$$\limsup_n |\frac{1}{b_n}\sum_{k=1}^n x_k| \leq \sup_{j,k\geq p} |z_j - z_k|,$$

quantité arbitrairement petite vu que z_n converge. $\diamond$.

Proposition 6.3.5. . *Soient $X_1, \ldots, X_n, \ldots$ une suite de v.a. réelles indépendantes et de carré intégrable et $b_n \uparrow_n +\infty$. On pose $S_n = X_1 + \ldots + X_n$. Alors, si $\sum_n \frac{Var(X_k)}{b_k^2} < +\infty$ et si $\frac{1}{b_n}\sum_{k=1}^n \mathbb{E}(X_k) \to_n m$, $\frac{S_n}{b_n} \to_n m$ p.s. et dans L^2.*

Preuve: On peut supposer les X_n centrées et alors $m = 0$. Vu le th. 6.3.3, $\sum_{k=1}^n \frac{X_k}{b_k}$ converge p.s. et donc (lem. 6.3.4) $\frac{S_n}{b_n} \to_n 0$ p.s. Quant à la convergence L^2, on a

$$\mathbb{E}(\frac{S_n^2}{b_n^2}) = \frac{1}{b_n^2}\sum_{k=1}^n \mathbb{E}(X_k^2) \to_n 0$$

puisque $\sum_n \frac{1}{b_k^2}\mathbb{E}(X_k^2)$ converge (lem. 6.3.4 pour la suite b_n^2). $\diamond$

Corollaire 6.3.6. . *Soient $X_1, \ldots, X_n, \ldots$ une suite de v.a. réelles indépendantes et de même loi avec $\mathbb{E}(X_1^2) < +\infty$. Alors $\frac{S_n}{n} \to_n \mathbb{E}(X_1)$ p.s. et dans L^2.*

Preuve: Il suffit de remarquer que $\sum_n \frac{\mathrm{Var}(X_n)}{n^2} = \mathrm{Var}(X_1)\sum_n \frac{1}{n^2} < +\infty$ et d'appliquer le th. 6.3.3. $\diamond$

Le cor. 6.3.6 établit la loi des grands nombres lorsque X_1 a un moment d'ordre deux fini.

6.4. La loi des grands nombres

6.4.1. On démontre la loi des grands nombres dans le cadre général.

Théorème 6.4.1. . *Soit $X_1, \dots, X_n, \dots$ une suite de v.a. réelles indépendantes et de même loi. On pose $S_n = X_1 + \dots + X_n$.*
(i) Si $\mathbb{E}(|X_1|) < +\infty$, $\frac{S_n}{n}$ converge p.s. et dans L^1 vers $\mathbb{E}(X_1)$.
(ii) Si $\frac{S_n}{n}$ converge p.s., $\mathbb{E}(|X_1|) < +\infty$.

D'abord deux lemmes relatifs à X v.a. réelle.

Lemme 6.4.2. . *On a $\sum_{n\geq 1} \mathbb{P}(|X| \geq n) \leq \mathbb{E}(|X|) \leq 1 + \sum_{n\geq 1} \mathbb{P}(|X| \geq n)$.*

Preuve: Soit $\phi(x) = \sum_{n\geq 1} 1_{\{x\geq n\}}$. On a, pour $x \in \overline{\mathbb{R}^+}$, $\phi(x) \leq x \leq 1 + \phi(x)$. D'où

$$\sum_{n\geq 1} \mathbb{P}(|X| \geq n) = \mathbb{E}(\sum_{n\geq 1} 1_{\{|X|\geq n\}}) \leq \mathbb{E}(|X|) \leq 1 + \mathbb{E}(\sum_{n\geq 1} 1_{\{|X|\geq n\}}) = 1 + \sum_{n\geq 1} \mathbb{P}(|X| \geq n). \diamond$$

Lemme 6.4.3. *On a $\sum_{n\geq 1} \mathbb{E}(\frac{X^2}{n^2} 1_{\{|X|<n\}}) \leq 2 + \mathbb{E}(|X|)$.*

Preuve: Vu que

$$k^2 \sum_{n=k}^{\infty} \frac{1}{n^2} = 1 + k^2 \sum_{n=k+1}^{\infty} \frac{1}{n^2} \leq 1 + k^2 \int_k^{\infty} \frac{1}{x^2}\, dx = 1 + k,$$

on a, tout étant positif,

$$\begin{aligned}
\sum_{n=1}^{\infty} \mathbb{E}(\frac{X^2}{n^2} 1_{\{|X|<n\}}) &= \sum_{n=1}^{\infty} \mathbb{E}(\frac{X^2}{n^2} \sum_{k=1}^{n} 1_{\{k-1\leq|X|<k\}}) = \sum_{k=1}^{\infty} \mathbb{E}(1_{\{k-1\leq|X|<k\}} X^2 \sum_{n=k}^{\infty} \frac{1}{n^2}) \\
&\leq \sum_{k=1}^{\infty} \mathbb{E}(1_{\{k-1\leq|X|<k\}} k^2 \sum_{n=k}^{\infty} \frac{1}{n^2}) \leq \sum_{k=1}^{\infty} \mathbb{E}(1_{\{k-1\leq|X|<k\}}(1+k)) \\
&\leq \sum_{k=1}^{\infty} \mathbb{E}(1_{\{k-1\leq|X|<k\}}(2+|X|)) \leq 2 + \mathbb{E}(|X|). \diamond
\end{aligned}$$

Revenons à la démonstration du théorème.

(i) On suppose $\mathbb{E}(|X_1|) < +\infty$. Posons $\hat{X}_k = X_k 1_{\{|X_k|<k\}}$, $\hat{S}_n = \sum_{k=1}^{n} \hat{X}_k$. Alors, vu le lem. 6.4.2,

$$\sum_k \mathbb{P}(X_k \neq \hat{X}_k) = \sum_k \mathbb{P}(|X_k| \geq k) = \sum_k \mathbb{P}(|X_1| \geq k) \leq \mathbb{E}(|X_1|) < +\infty.$$

Donc (Borel-Cantelli) $X_k = \hat{X}_k$ à partir d'un certain rang p.s et $\frac{S_n}{n} - \frac{\hat{S}_n}{n} \to_n 0$ p.s. On est donc ramené à étudier la limite de $\frac{\hat{S}_n}{n}$. Pour cela, on utilise la prop. 6.3.5.

D'une part, vu le lem. 6.4.3,

$$\sum_{n\geq 1}\frac{\mathrm{Var}(\hat{X}_n)}{n^2}\leq \sum_{n\geq 1}\frac{\mathbb{E}(\hat{X}_n^2)}{n^2}=\sum_{n\geq 1}\mathbb{E}(\frac{X_1^2}{n^2}1_{\{|X_1|<n\}})\leq 2+\mathbb{E}(|X_1|)<+\infty.$$

D'autre part, comme $\mathbb{E}(\hat{X}_k)=\mathbb{E}(X_k 1_{\{|X_k|<k\}})=\mathbb{E}(X_1 1_{\{|X_1|<k\}})\to_k \mathbb{E}(X_1)$ (Lebesgue), $\frac{1}{n}\mathbb{E}(\hat{S}_n)\to_n \mathbb{E}(X_1)$. Finalement $\frac{\hat{S}_n}{n}\to_n \mathbb{E}(X_1)$ p.s. et il en est de même de $\frac{S_n}{n}$.

Passons à la convergence dans L^1. On peut supposer $\mathbb{E}(X_1)=0$. On a, pour tout $M>0$,

$$\mathbb{E}(|\frac{S_n}{n}|)\leq \mathbb{E}(|\frac{1}{n}\sum_{k=1}^n X_k 1_{\{|X_k|<M\}}|)+\mathbb{E}(|\frac{1}{n}\sum_{k=1}^n X_k 1_{\{|X_k|\geq M\}}|).$$

D'une part, vu la première partie et que $0=\mathbb{E}(X_1)=\mathbb{E}(X_1 1_{\{|X_1|<M\}})+\mathbb{E}(X_1 1_{\{|X_1|\geq M\}})$,

$$|\frac{1}{n}\sum_{k=1}^n X_k 1_{\{|X_k|<M\}}|\to_n |\mathbb{E}(X_1 1_{\{|X_1|<M\}})|=|\mathbb{E}(X_1 1_{\{|X_1|\geq M\}})|$$

p.s. en restant borné par M et donc aussi dans L^1. D'autre part

$$\mathbb{E}(|\frac{1}{n}\sum_{k=1}^n X_k 1_{\{|X_k|\geq M\}}|)=\frac{1}{n}\mathbb{E}(|\sum_{k=1}^n X_1 1_{\{|X_1|\geq M\}}|)\leq \mathbb{E}(|X_1|1_{\{|X_1|\geq M\}}).$$

. D'où

$$\limsup_n \mathbb{E}(|\frac{S_n}{n}|)\leq |\mathbb{E}(X_1 1_{\{|X_1|\geq M\}})|+\mathbb{E}(|X_1|1_{\{|X_1|\geq M\}})\leq 2\mathbb{E}(|X_1|1_{\{|X_1|\geq M\}}).$$

Mais cette dernière quantité est arbitrairement petite puisque $\mathbb{E}(|X_1|1_{\{|X_1|\geq M\}})\to 0$ lorsque $M\to+\infty$ (Lebesgue).

(ii) Supposons que $\frac{S_n}{n}$ converge p.s. Donc (cor. 6.2.2) $\frac{S_n}{n}\to_n c$ p.s et $\frac{X_n}{n}=\frac{S_n}{n}-\frac{n-1}{n}\frac{S_{n-1}}{n-1}\to 0$ p.s. Ceci implique que $\mathbb{P}(\limsup\{|X_n|\geq n\})=0$ et donc (prop. 4.1.2) que $\sum_n \mathbb{P}(|X_n|\geq n)<+\infty$. On a alors (lem. 6.4.2)

$$\mathbb{E}(|X_1|)\leq 1+\sum_n \mathbb{P}(|X_1|\geq n)=1+\sum_n \mathbb{P}(|X_n|\geq n)<+\infty \quad \diamond.$$

<u>Remarque 1</u>. Tradtionnellement le th. 6.4.1 s'appelle la loi forte des grands nombres. On réserve le nom de loi faible des grands nombres à la convergence en probabilité de S_n/n vers $\mathbb{E}(X_1)$ qui est évidemment une conséquence de la loi forte.

<u>Remarque 2</u>. Soit μ une probabilité sur un espace mesurable $(E,\mathcal{E})$. Le tirage d'une suite de points de E selon μ peut se représenter par une suite de v.a. indépendantes

de loi μ. Soit $A \in \mathcal{E}$. Les v.a. $1_A(X_1), 1_A(X_2), \ldots, 1_A(X_n), \ldots$ sont indépendantes, de même loi, d'espérance $\mu(A)$. On a donc p.s.

$$\mu(A) = \lim_n \frac{1}{n} \sum_{k=1}^{n} 1_A(X_k) = \lim_n \frac{\text{nombre de } k \leq n \text{ tels que } X_k \in A}{n}.$$

On retrouve là la justification fréquentielle de la notion de probabilités.

Remarque 2. En raisonnant composante par composante, le th. 6.4.1 se généralise immédiatement aux v.a. à valeurs $\mathbb{R}^d$.

6.4.2. Nombres au hasard. On revient sur la question, posée en 4.8.1, de construire une suite $(u_n,\, n \geq 1)$ de nombres compris entre 0 et 1 et pouvant être considérée comme le résultat de tirages indépendants selon la loi $U(0,1)$. Soit $(U_n,\, n \geq 1)$ une suite de v.a. indépendantes de loi $U(0,1)$. On a (loi des grands nombres), pour tous $0 \leq a < b \leq 1$,

$$\frac{1}{n} \sum_{k=1}^{n} 1_{[a,b]}(U_k) \to_n b - a \quad \text{p.s.}$$

Mais $X_1 = (U_1, U_2), X_2 = (U_3, U_4), \ldots, X_n = (U_{2n-1}, U_{2n}), \ldots$ est aussi une suite de v.a. indépendantes à valeurs $\mathbb{R}^2$ de loi uniforme sur $[0,1] \times [0,1]$ et l'on a, pour tous $0 \leq a_1 < b_1 \leq 1, 0 \leq a_2 < b_2 \leq 1$, posant $D = [a_1, b_1] \times [a_2, b_2]$

$$\frac{1}{n} \sum_{j=0}^{n} 1_D(U_{2j+1}, U_{2j+2}) \to_n (b_1 - a_1)(b_2 - a_2) \ \text{ p.s}$$

Plus généralement, pour tout k et tous $0 \leq a_1 < b_1 \leq 1, \ldots, 0 \leq a_k < b_k \leq 1$, posant $D = \prod_{j=1}^{k} [a_j, b_j]$,

$$\frac{1}{n} \sum_{j=0}^{n} 1_D(U_{kj+1}, \ldots, U_{kj+k}) \to_n \prod_{j=1}^{k} (b_j - a_j) \ \text{ p.s}$$

Ceci conduit à:

Définition 6.4.4. *Une suite $(u_n,\, n \geq 1)$ de nombres compris entre 0 et 1 est dite k-uniforme ($k \in \mathbb{N}^*$) si, pour tous $0 \leq a_1 < b_1 \leq 1, \ldots, 0 \leq a_k < b_k \leq 1$, posant $D = \prod_{j=1}^{k} [a_j, b_j]$,*

$$\frac{1}{n} \sum_{j=0}^{n} 1_D(u_{kj+1}, \ldots, u_{kj+k}) \to_n \prod_{j=1}^{k} (b_j - a_j).$$

L'idéal pour qu'une suite $(u_n,\, n \geq 1)$ puisse être considérée comme le résultat de tirages indépendants selon la loi uniforme sur $[0,1]$ serait que cette suite soit k-uniforme pour tout k mais ceci, en pratique, est impossible et on se contente d'approximations.

On utilise fréquemment des algorithmes du type suivant. On choisit $M \in \mathbb{N}$ grand (de l'ordre de 10^8) et une application g de $E = \{0, 1 \ldots, M-1\}$ dans lui-même. On se donne $v_0 \in E$ et on pose $v_{n+1} = g(v_n)$, $u_n = v_n/M$. Les différents choix de v_0 engendrent différentes suites. Une telle suite étant nécessairement périodique, ceci n'est qu'une approximation. On peut prendre $M = 2^{31}$ et $g(x) = 7^5 x$ modulo M.

6.4.3. Méthode de Monte-Carlo. Le principe de la méthode est le suivant. Soient f une densité sur $\mathbb{R}^d$, $(X_n,\, n \geq 1)$ une suite de v.a. indépendantes de densité f et $\phi \in L^1(f.\lambda)$. Alors, d'après la loi des grands nombres,

$$I_n = \frac{1}{n}\sum_{k=1}^{n} \phi(X_k) \to_n \mathbb{E}(\phi(X_1)) = \int \phi(x) f(x)\, dx = I \quad \text{p.s.}$$

Donc, si on sait simuler des v.a. de densité f, on peut obtenir une valeur approchée de I. Noter que I_n se met sous forme récursive:

$$I_{n+1} = I_n + \frac{1}{n+1}(\phi(X_{n+1}) - I_n),$$

ce qui rend le calcul agréable. Examinons de plus près deux cas.

1. On veut calculer $\int_D h(x)\, dx$, D étant un domaine borné de $\mathbb{R}^d$ et $h1_D$ intégrable. Soient $\Delta = \prod_{k=1}^d [a_k, b_k] \supset D$, $V = \prod_{k=1}^d (b_k - a_k)$ et $(X_n,\, n \geq 1)$ une une suite de v.a. indépendantes de loi uniforme sur Δ. On peut appliquer le résultat précédent à $f = \frac{1}{V} 1_\Delta$, $\phi = h1_D$ et on a

$$\frac{V}{n}\sum_{k=1}^{n} h(X_k) 1_D(X_k) \to_n V \frac{1}{V}\int h(x) 1_D(x) 1_\Delta(x)\, dx = \int_D h(x)\, dx \quad \text{p.s.}$$

2. On veut calculer $\int \phi(x) f(x)\, dx$ (f densité et $\phi \in L^1(f.\lambda)$) et on sait simuler des v.a. $(Y_n,\, n \geq 1)$ indépendantes de densité g avec $f \leq a\, g$. Alors on peut utiliser la prop. 4.8.3 pour simuler des v.a. de densité f mais, en fait, on a directement:

Proposition 6.4.5. *Soient f, g deux densités sur $\mathbb{R}^d$ telles que $f \leq a\, g$, $(Y_n,\, n \geq 1)$ et $(U_n,\, n \geq 1)$ deux suites de v.a. indépendantes de lois respectives $g.\lambda$ et $U(0,1)$ et indépendantes entre elles. Alors, pour toute $\phi \in L^1(f.\lambda)$,*

$$\frac{a}{n}\sum_{k=1}^{n} \phi(Y_k) 1_{\{aU_k g(Y_k) < f(Y_k)\}} \to_n \int_{\mathbb{R}^d} \phi(x) f(x)\, dx \quad \textit{p.s.}$$

Preuve: Les v.a. $(\phi(Y_k) 1_{\{aU_k g(Y_k) < f(Y_k)\}},\, k \geq 1)$ étant indépendantes, il suffit d'appliquer la loi des grands nombres vu que $\mathbb{E}(\phi(Y_1) 1_{\{aU_1 g(Y_1) < f(Y_1)\}}) = \frac{1}{a}\int \phi f\, d\lambda$ pour $\phi \geq 0$ (lem. 4.8.4) puis, par différence, pour $\phi \in L^1(f.\lambda)$. $\diamond$

Pour être complet, il faudrait considérer les vitesses de convergence. (On dit que a_n converge vers a à la vitesse $\frac{1}{n^\alpha}$ si $|a - a_n| = O(\frac{1}{n^\alpha})$). Vu le th.7.3.1 qu'on verra au chapitre suivant, cette vitesse est, en général, de l'ordre de $\frac{1}{\sqrt{n}}$ ce qui fait que, pour des petites valeurs de d, cette méthode est peu compétitive par rapport aux méthodes classiques d'analyse numérique mais que, pour des valeurs assez grandes de d, elle devient intéressante.

6.5. Complément: critère des trois séries.

6.5.1. On examine la réciproque du th. 6.3.3.

Proposition 6.5.1. *Soit $X_1, \dots, X_n, \dots$ une suite de v.a.r. indépendantes. On suppose qu'il existe $M > 0$ tel que, pour tout n, $|X_n| \le M$ p.s. Alors, si $S_n = \sum_{k=1}^n X_k$ converge p.s., les séries $\sum \mathbb{E}(X_n)$ et $\sum$ Var(X_n) sont convergentes.*

Preuve: Elle repose sur le lemme suivant.

Lemme 6.5.2. *Soit X une v.a.r. centrée vérifiant $|X| \le M$ p.s. On pose $\sigma^2 = E(X^2)$ et on note $\phi(t)$ sa fonction caractéristique. alors, si $|t| \le M^{-1}$,*

$$|\phi(t)| \le \exp(-\frac{1}{3}\sigma^2 t^2).$$

Preuve: Puisque $E(|X|^3) < +\infty$, on a $\phi^{(3)}(t) = (i)^3 E[X^3 e^{itX}]$ et $|\phi^{(3)}(t)| \le E(|X|^3) \le M\sigma^2$. Vu que $\phi'(0) = 0$ et $\phi''(0) = -\sigma^2$, on a $\phi(t) = 1 - \frac{\sigma^2}{2}t^2 + r(t)$ avec $|r(t)| \le \frac{|t|^3}{6} \sup_{ls| \le |t|} |\phi^{(3)}(s)| \le \frac{|t|^3}{6} M\sigma^2$. Alors, si $|t| \le M^{-1}$, $\sigma^2 t^2 \le \sigma^2 M^{-2} \le 1$ et

$$|\phi(t)| \le 1 - \frac{\sigma^2}{2}t^2 + \frac{|t|^3}{6}M\sigma^2 \le 1 - \frac{\sigma^2}{2}t^2 + \frac{t^2}{6}\sigma^2 = 1 - \frac{\sigma^2}{3}t^2 \le \exp(-\frac{1}{3}\sigma^2 t^2). \diamond$$

On pose $Y_k = X_k - E(X_k)$ (alors $|Y_k| \le 2M$ p.s.), $\sigma_k^2 = \text{Var}(X_k)$, $S_n = X_1 + ... + X_n$, $\overline{S}_n = Y_1 + ... + Y_n$ et on note ϕ_n et ψ_n les fonctions caractéristiques de S_n et $\overline{S}_n$. Puisque $\phi_n(t) = \exp(itE(S_n))\psi_n(t)$, on a, d'après le lem. 6.5.2, pour $|t| \le (2M)^{-1}$,

$$|\phi_n(t)| = |\psi_n(t)| = \prod_{k=1}^n |\phi_{Y_k}(t)| \le \exp(-\frac{1}{3}t^2 \sum_{k=1}^n \sigma_k^2).$$

Supposons que $\sum_k \sigma_k^2 = +\infty$. Alors, pour tout t tel que $|t| \le (2M)^{-1}$, $|\phi_n(t)| \to 1_{\{0\}}(t)$. Mais, par hypothèse, S_n converge vers S p. s. et donc (Lebesgue) $\phi_{S_n}(t) \to \phi_S(t)$ d'où , pour tout t, $|\phi_n(t)| \to |\phi_S(t)|$ qui est continue. On a donc $\sum_k \sigma_k^2 < +\infty$. Comme $\sigma_n^2 = \text{Var}(X_n) = \text{Var}(Y_n)$ et que $E(Y_n) = 0$, $\sum_k \sigma_k^2 < +\infty$ implique (th. 6.3.3) que $\sum Y_n$ converge p.s. Mais alors $\sum X_n$ et $\sum(X_n - E(X_n))$ convergent p.s. donc, par différence, $\sum E(X_n)$ converge. $\diamond$

6.5.2. Critère des trois séries.

Théorème 6.5.3. *Soient $X_1, \dots, X_n, \dots$ une suite de v.a. réelles indépendantes et $K > 0$. On pose $X_n^K = X_n 1_{\{|X_n| \le K\}}$. Il y a équivalence entre*
(i) $\sum_{k=1}^n X_k$ converge p.s.
(ii) Les séries $\sum \mathbb{P}(|X_n| > K)$, $\sum \mathbb{E}(X_n^K)$, $\sum$ Var(X_n^K) convergent.

Preuve: (i) Supposons que $\sum_n X_n$ converge p.s. Alors $\sum \mathbb{P}(|X_n| > K) < +\infty$ car, si $\sum \mathbb{P}(|X_n| > K) = +\infty$, on a p.s. (prop. 4.1.2) $|X_n| > K$ infiniment souvent et S_n diverge p.s. Alors la convergence de $\sum \mathbb{P}(|X_n| > K)$ implique (prop. 4.1.2) que p.s.

$|X_n| > K$ n'a lieu qu'un nombre fini de fois. Les séries $\sum X_n$ et $\sum X_n^K$ sont donc p.s. de même nature et $\sum X_n^K$ converge p.s. Puisque $|X_n^K| \le K$, on peut appliquer la prop. 6.5.1 et $\sum E(X_n^K)$ et $\sum \mathrm{Var}(X_n^K)$ convergent.

(ii) Supposons que les trois séries convergent. Vu la prop. 6.5.1, $\sum X_n^K$ converge p.s. et, comme ci-dessus, la convergence de $\sum \mathbb{P}(|X_n| > K)$ implique que les séries $\sum X_n$ et $\sum X_n^K$ sont p.s. de même nature. Donc $\sum_n X_n$ converge p.s. $\diamond$

6.6. Complément: grandes déviations.

6.6.1. Soit $X_1, \dots, X_n, \dots$ une suite de v.a.r. indépendantes et de même loi μ avec $\mathbb{E}|X_1| < +\infty$. On pose $m = \mathbb{E}(X_1)$. Si $a > m$, il résulte du th. 6.4.1 que, posant $S_n = X_1 + \ldots + X_n$,

$$\mathbb{P}(\frac{S_n}{n} > a) \to_n 0.$$

On voudrait préciser la vitesse de convergence. On sait que plus une v.a.r. posséde de moments finis, plus on peut espérer des estimatuions précises. On pose donc:

$$\phi(\lambda) = \mathbb{E}(\mathrm{e}^{\lambda X_1}) = \int \mathrm{e}^{\lambda x}\, d\mu(x), \quad G(\lambda) = \log \phi(\lambda), \quad \Delta = \{\lambda,\, \phi(\lambda) < +\infty\} \tag{6.2}$$

et **on suppose que 0 est un point intérieur de Δ** . La fonction $\phi(\lambda)$ est strictement positive et, vu que

$$\forall a < b < c < d,\ \forall n \ge 0,\ \exists M\ \forall \lambda \in [b, c],\ |x^n\, \mathrm{e}^{\lambda x}| \le M\,(\mathrm{e}^{ax} + \mathrm{e}^{dx}),$$

Δ est un intervalle, ϕ est indéfiniment dérivable sur $\overset{\circ}{\Delta}$ et $\phi^{(n)}(\lambda) = \int x^n\, \mathrm{e}^{\lambda x}\, d\mu(x)$ d'après la prop. 3.3.7. En particulier $\phi(0) = 1$, $\phi'(0) = \int x\, d\mu(x) = m$.

La fonction ϕ étant strictement positive, G est aussi indéfiniment dérivable sur $\overset{\circ}{\Delta}$ et l'on a, pour $\lambda \in \overset{\circ}{\Delta}$,

$$G(0) = 0,\ \ G'(\lambda) = \frac{\phi'(\lambda)}{\phi(\lambda)} = \int y\, \mathrm{e}^{\lambda y - G(\lambda)}\, d\mu(y),\ \ G'(0) = m.$$

Enfin G est convexe puisque, pour $0 < \alpha < 1$, vu l'inégalité de Hölder:

$$\phi(\alpha\lambda_1 + (1-\alpha)\lambda_2) = \int \mathrm{e}^{\alpha\lambda_1 x}\mathrm{e}^{(1-\alpha)\lambda_2 x}\, d\mu(x) \le [\int \mathrm{e}^{\lambda_1 x}\, d\mu(x)]^{\alpha}[\int \mathrm{e}^{\lambda_2 x}\, d\mu(x)]^{1-\alpha},$$

$$G(\alpha\lambda_1 + (1-\alpha)\lambda_2) \le \alpha \log \phi(\lambda_1) + (1-\alpha)\log\phi(\lambda_2 x) = \alpha G(\lambda_1) + (1-\alpha)G(\lambda_2 x).$$

6.6.2. Majoration. On a alors, pour $a > m$ et tout $\lambda > 0$,

$$\mathbb{P}(\frac{S_n}{n} \ge a) = \mathbb{P}(\mathrm{e}^{\lambda S_n} \ge \mathrm{e}^{\lambda n a}) \le \mathrm{e}^{-\lambda n a}\mathbb{E}(\mathrm{e}^{\lambda S_n}) = \mathrm{e}^{-\lambda n a}\,[\phi(\lambda)]^n = \exp(-n(\lambda a - G(\lambda)))$$

d'où

$$\mathbb{P}(\frac{S_n}{n} \ge a) \le \exp(-n \sup_{\lambda > 0}(\lambda a - G(\lambda))). \tag{6.3}$$

Ceci conduit à s'intéresser à la fonction

$$I(x) = \sup_{\lambda\in\mathbb{R}}(\lambda x - G(\lambda)), \quad x \in \mathbb{R}. \tag{6.4}$$

Cette fonction s'appelle la transformée de Legendre de G. Elle joue un rôle important en analyse convexe. Indiquons quelques propriétés.

Lemme 6.6.1. *La fonction $I(x)$ est positive, convexe, vérifie $I(m) = 0$, est décroissante sur $]-\infty, m]$ et croissante sur $[m, +\infty[$. Pour $x > m$, $I(x) = \sup_{\lambda>0}(\lambda x - G(\lambda))$.*

Preuve: Vu que, pour $\lambda = 0$, $\lambda x - G(\lambda) = 0$, $I(x) \geq 0$. La fonction I étant un sup de fonctions affines, elle est convexe. De plus, d'après l'inégalité de Jensen,

$$e^{G(\lambda)} = \mathbb{E}(e^{\lambda X_1}) \geq e^{\lambda\mathbb{E}(X_1)} = e^{\lambda m},$$

d'où, pour tout λ, $\lambda m \leq G(\lambda)$ et donc $I(m) \leq 0$ et $I(m) = 0$. De plus la fonction I étant positive, convexe et nulle en m, elle croit sur $[m, +\infty[$ et décroit sur $]-\infty, m]$. Enfin la fonction $h(\lambda) = \lambda x - G(\lambda)$ est concave, dérivable au voisinage de 0 et vérifie $h(0) = 0$, $h'(0) = x - G'(0) = x - m > 0$ et donc $\sup_{\lambda>0}(\lambda x - G(\lambda)) = \sup_{\lambda\in\mathbb{R}}(\lambda x - G(\lambda))$. ⋄

On en déduit immédiatement les inégalités de Chernov:

Proposition 6.6.2. *On a:*
(i) pour tout $a \geq m$, ,$\mathbb{P}(\frac{S_n}{n} \geq a) \leq e^{-nI(a)}$,
(ii) pour tout $a \leq m$, $\mathbb{P}(\frac{S_n}{n} \leq a) \leq e^{-nI(a)}$.

Preuve: (i) résulte de (6.4) et du lem. 6.6.1 pour $a > m$ et est évident pour $a = m$ puisque $I(m) = 0$. (ii) s'obtient en appliquant (i) à la suite $(-X_n)$. ⋄

6.6.3. Minoration.

Proposition 6.6.3. *On a, pour tous $a \in \mathbb{R}$ et $\delta > 0$,*

$$\liminf_n \frac{1}{n} \log \mathbb{P}(|\frac{S_n}{n} - a| < \delta) \geq -I(a).$$

Preuve: Si $I(a) = +\infty$, il n'y a rien à montrer. On suppose donc $I(a) < +\infty$. La preuve repose sur l'étude de plusieurs cas selon que $h(\lambda) = \lambda a - G(\lambda)$ atteint son maximum ou non.

(i) On suppose qu'il existe $\lambda_0 \in \overset{\circ}{\Delta}$ tel que $I(a) = \lambda_0 a - G(\lambda_0)$. La fonction h étant dérivable sur $\overset{\circ}{\Delta}$, on a $h'(\lambda_0) = 0$ i.e. $G'(\lambda_0) = a$. Soient ν la probabilité sur $\mathbb{R}$ définie par:

$$d\nu(x) = \phi^{-1}(\lambda_0)\, e^{\lambda_0 x}\, d\mu(x) \tag{6.5}$$

et $Y_1, \ldots, Y_n, \ldots$ une suite de v.a.r. indépendantes de loi ν définies sur $(\Omega', \mathcal{A}', \mathbb{P}')$. On pose $\Sigma_n = Y_1 + \ldots + Y_n$. On vérifie facilement que, notant $\mathbb{E}'(Z)$ pour $\int Z\, d\mathbb{P}'$, $\mathbb{E}'(|Y_1|) =< +\infty$ et que

$$\mathbb{E}'(Y_1) = \int x\, d\nu(x) = \phi^{-1}(\lambda_0) \int x\, e^{\lambda_0 x}\, d\mu(x) = \frac{\phi'(\lambda_0)}{\phi(\lambda_0)} = G'(\lambda_0) = a.$$

D'autre part, pour toute $f \in \mathcal{B}^+(\mathbb{R})$,

$$\mathbb{E}(f(S_n)) = \int f(x_1,\ldots,x_n)\,d\mu(x_1)\ldots d\mu(x_n)$$
$$= \phi^n(\lambda_0)\int f(x_1,\ldots,x_n)\,\mathrm{e}^{-\lambda_0(x_1+\ldots+x_n)}\,d\nu(x_1)\ldots d\nu(x_n) = \phi^n(\lambda_0)\mathbb{E}'(f(\Sigma_n)\mathrm{e}^{-\lambda_0\Sigma_n}).$$

On en déduit que, pour tout $\varepsilon\in]0,\delta]$,

$$\mathbb{P}(|\frac{S_n}{n}-a|<\delta)\ge\mathbb{P}(|\frac{S_n}{n}-a|<\varepsilon)=\phi^n(\lambda_0)\mathbb{E}'(1_{\{|\frac{\Sigma_n}{n}-a|<\varepsilon\}}\mathrm{e}^{-\lambda_0\Sigma_n})$$
$$=\phi^n(\lambda_0)\mathrm{e}^{-na\lambda_0}\mathbb{E}'(1_{\{|\frac{\Sigma_n}{n}-a|<\varepsilon\}}\mathrm{e}^{-\lambda_0 n(\frac{\Sigma_n}{n}-a))})\ge\phi^n(\lambda_0)\mathrm{e}^{-na\lambda_0}\mathrm{e}^{-n\varepsilon\lambda_0}\mathbb{P}'(|\frac{\Sigma_n}{n}-a|<\varepsilon).$$

D'où

$$\frac{1}{n}\log\mathbb{P}(|\frac{S_n}{n}-a|<\delta)\ge -a\lambda_0+G(\lambda_0)-\lambda_0\varepsilon+\frac{1}{n}\log\mathbb{P}'(|\frac{\Sigma_n}{n}-a|<\varepsilon)$$

et, puisque $-a\lambda_0+G(\lambda_0)=-I(a)$ et que $\mathbb{P}'(|\frac{\Sigma_n}{n}-a|<\varepsilon)\to_n 1$ (loi des grands nombres),

$$\liminf_n\frac{1}{n}\log\mathbb{P}(|\frac{S_n}{n}-a|<\delta)\ge -I(a)-\lambda_0\varepsilon.$$

Ce qui établit la proposition dans ce cas.

(ii) On suppose qu'il existe $\lambda_k\in\overset{\circ}{\Delta}$, $\lambda_k\uparrow+\infty$, tels que $I(a)=\lim_k\lambda_k a-G(\lambda_k)$. On a alors

$$\mathrm{e}^{-I(a)}=\lim_k\mathrm{e}^{G(\lambda_k)-\lambda_k a}=\lim_k\int\mathrm{e}^{\lambda_k(x-a)}\,d\mu(x).$$

Puisque $\int_{]-\infty,a[}\mathrm{e}^{\lambda_k(x-a)}\,d\mu(x)\to_k 0$, $\int_{[a,+\infty[}\mathrm{e}^{\lambda_k(x-a)}\,d\mu(x)\to_k\mathrm{e}^{-I(a)}$ ce qui implique, vu que $\mathrm{e}^{\lambda_k(x-a)}\uparrow+\infty$ sur $]a,+\infty]$, que $\mu(]a,+\infty[)=0$ et donc que $\mathrm{e}^{-I(a)}=\mu(\{a\})$. Alors

$$\mathbb{P}(|\frac{S_n}{n}-a|<\delta)\ge\mathbb{P}(X_1=\ldots=X_n=a)=[\mu(\{a\})]^n=\mathrm{e}^{-nI(a)}$$

et la minoration cherchée.

Supposons:

$$\text{pour tout } \lambda\in\mathbb{R},\quad\int e^{\lambda x}\,d\mu(x)=\mathbb{E}(\mathrm{e}^{\lambda X_1})<+\infty. \qquad (6.6)$$

Alors $\Delta=\mathbb{R}$, $G(\lambda)$ est partout finie et $h(\lambda)=\lambda a-G(\lambda)$ est une fonction concave C^∞ sur $\mathbb{R}$ et on est nécessaiement soit dans le cas (i), soit dans le cas (ii), ce qui prouve la proposition sous cette hypothèse.

Une autre situation intéressante est la suivante. Rappelons que le support S_μ de μ est le plus petit fermé F tel que $\mu(F^c)=0$. On pose $\alpha_\mu=\inf S_\mu$, $\beta_\mu=\sup S_\mu$ (les valeurs infinies ne sont pas exclues). Considérons l'hypothèse:

$$\text{pour tout } a\in]\alpha_\mu,\beta_\mu[,\ \text{ il existe } \lambda\in\overset{\circ}{\Delta}\ \text{ tel que } G'(\lambda)=a. \qquad (6.7)$$

Si $a \in]\alpha_\mu, \beta_\mu[$, on est dans la cas (i). Supposons $\beta_\mu < +\infty$ et $a \geq \beta_\mu$. On a

$$\text{pour tout } \lambda > 0, \quad \int \mathrm{e}^{\lambda x}\, d\mu(x) = \int_{]-\infty,\beta_\mu]} \mathrm{e}^{\lambda x}\, d\mu(x) \leq \mathrm{e}^{\lambda\beta_\mu} < +\infty$$

ce qui implique que $\mathbb{R}^+ \subset \Delta$. Mais, sur $\overset{\circ}{\Delta}$,

$$G'(\lambda) = \int_{]-\infty,\beta_\mu]} x\, \mathrm{e}^{\lambda x}\, d\mu(x) \,/ \int_{]-\infty,\beta_\mu]} \mathrm{e}^{\lambda x}\, d\mu(x) \leq \beta_\mu \leq a.$$

La fonction $h(\lambda) = \lambda a - G(\lambda)$ est donc croissante sur $\Delta \supset \mathbb{R}^+$ (h' est ≥ 0) et on a

$$I(a) = \sup_{\lambda\in\Delta}(\lambda a - G(\lambda)) = \lim_{\lambda\to+\infty}(\lambda a - G(\lambda)).$$

On est dans le cas (ii). (Noter que, si $a > \beta_\mu$, $I(a) = +\infty$ puisque $\mu(\{a\})) = 0$). Enfin on a le même résultat pour $a \leq \alpha_\mu$ en considérant la suite $(-X_n)$, ce qui montre la proposition sous l'hypothèse (6.7). Il reste à examiner quelques situations spéciales que nous admettons. $\diamond$

6.6.4. Le théorème de Cramer.

Théorème 6.6.4. *Soit $X_1, \ldots, X_n, \ldots$ une suite de v.a.r. indépendantes et de même loi μ. On suppose que $\int e^{\lambda x}\, d\mu(x) < +\infty$ si $|\lambda| \leq \lambda_0$, $\lambda_0 > 0$. On pose:*

$$S_n = X_1 + \ldots + X_n, \quad G(\lambda) = \log \int e^{\lambda x}\, d\mu(x), \quad I(x) = \sup_{\lambda\in\mathbb{R}}(\lambda x - G(\lambda)).$$

Alors on a:

$$\text{pour tout fermé } F \text{ de } \mathbb{R}, \quad \limsup_n \frac{1}{n} \log \mathbb{P}(\frac{S_n}{n} \in F) \leq -\inf_{x\in F} I(x),$$

$$\text{pour tout ouvert } G \text{ de } \mathbb{R}, \quad \liminf_n \frac{1}{n} \log \mathbb{P}(\frac{S_n}{n} \in G) \geq -\inf_{x\in G} I(x).$$

Preuve: La minoration est une conséquence immédiate de la prop. 6.6.3 car, si $x \in G$, il existe $\delta > 0$ tel que $\{y,\, |y-x| < \delta\} \subset G$ et $\mathbb{P}(\frac{S_n}{n} \in G) \geq \mathbb{P}(|\frac{S_n}{n} - x| < \delta)$. Passons à la majoration. Supposons que $F^+ = F \cap [m, +\infty[\neq \emptyset$ et $F^- = F \cap]-\infty, m] \neq \emptyset$. Soient $b^+ = \inf F \cap [m, +\infty[$ et $b^- = \sup F \cap]-\infty, m]$. On a, vu la prop. 6.6.2 et la monotonie de I sur $]-\infty, m]$ et $[m, +\infty[$,

$$\mathbb{P}(\frac{S_n}{n} \in F^+) \leq \mathbb{P}(\frac{S_n}{n} \geq b^+) \leq \exp(-nI(b^+)) \leq \exp(-n \inf_{x\in F} I(x)),$$
$$\mathbb{P}(\frac{S_n}{n} \in F^-) \leq \mathbb{P}(\frac{S_n}{n} \leq b^-) \leq \exp(-nI(b^-)) \leq \exp(-n \inf_{x\in F} I(x)),$$
$$\mathbb{P}(\frac{S_n}{n} \in F) \leq \mathbb{P}(\frac{S_n}{n} \in F^+) + \mathbb{P}(\frac{S_n}{n} \in F^-) \leq 2\exp(-n \inf_{x\in F} I(x)).$$

On conclut facilement puisque $\frac{1}{n}\log 2 \to_n 0$! Si $F^- = \emptyset$ (resp. $F^+ = \emptyset$), il suffit de considérer la majoration ci-dessus pour F^+ (resp. F^-). $\diamond$

Corollaire 6.6.5. *Sous les hypothèses du th. 6.6.4, si I est continue au point a,*

$$\text{si } a > m,\ \lim_n \frac{1}{n}\log\mathbb{P}(\frac{S_n}{n} \geq a) = -I(a),\quad \text{si } a < m,\ \lim_n \frac{1}{n}\log\mathbb{P}(\frac{S_n}{n} \leq a) = -I(a).$$

Preuve: Supposons $a > m$. D'une part $\limsup_n \frac{1}{n}\log\mathbb{P}(\frac{S_n}{n} \geq a) \leq -I(a)$ et d'autre part

$$\liminf_n \frac{1}{n}\log\mathbb{P}(\frac{S_n}{n} \geq a) \geq \liminf_n \frac{1}{n}\log\mathbb{P}(\frac{S_n}{n} > a) \geq -\inf_{x>a} I(x) = -I(a).\ \diamond$$

6.6.5. Exemples.

a. $\mu = \delta_m$ i.e. $\mathbb{P}(X_1 = m) = 1$. On a:

$$\Delta = \mathbb{R},\ \ \phi(\lambda) = \mathrm{e}^{\lambda m},\ \ G(\lambda) = \lambda m,$$
$$I(x) = 0 \text{ si } x = m,\ \ I(x) = +\infty \text{ si } x \neq m.$$

b. $\mu = p\delta_1 + (1-p)\delta_0$ $(0 < p < 1)$ i.e. $\mathbb{P}(X_1 = 1) = p$, $\mathbb{P}(X_1 = 0) = 1 - p$. On a:

$$\Delta = \mathbb{R},\ \ \phi(\lambda) = p\mathrm{e}^{\lambda} + 1 - p,\ \ G(\lambda) = \log(p\mathrm{e}^{\lambda} + 1 - p),$$
$$I(x) = x\log(\frac{x}{p}) + (1-x)\log(\frac{1-x}{1-p}) \text{ si } x \in [0,1],\ \ I(x) = +\infty \text{ si } x \notin [0,1].$$

c. $\mu = N_1(m, \sigma^2)$ i.e. $d\mu(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp(-\frac{1}{2\sigma^2}(x-m)^2)\,dx$. On a:

$$\Delta = \mathbb{R},\ \ \phi(\lambda) = \exp(m\lambda + \frac{\sigma^2\lambda^2}{2}),\ \ G(\lambda) = m\lambda + \frac{\sigma^2\lambda^2}{2},$$
$$I(x) = \frac{(x-m)^2}{2\sigma^2}.$$

c. $\mu = G(1, \gamma)$ i.e. $d\mu(x) = \gamma\mathrm{e}^{-\gamma x}1_{]0,+\infty[}(x)\,dx$. On a:

$$\Delta =]-\infty, \gamma[,\ \ \phi(\lambda) = \frac{\gamma}{\gamma-\lambda},\ \lambda < \gamma,\ \ G(\lambda) = \log(\frac{\gamma}{\gamma-\lambda}),\ \lambda < \gamma,\ G(x) = +\infty,\ \lambda \geq \gamma,$$
$$I(x) = \gamma x - 1 - \log(\gamma x) \text{ si } x > 0,\ \ I(x) = +\infty \text{ si } x \leq 0.$$

Noter que μ a pour support $[0, +\infty[$ et que, pour tout $a > 0$, l'équation $G'(\lambda) = a$ s'écrit $\frac{1}{\gamma-\lambda} = a$ et a pour solution $\lambda = \gamma - \frac{1}{a} \in]-\infty, \gamma[$. La condition (6.7) est bien vérifiée dans ce cas.

Chapitre 7

Convergence en loi

7.1. Convergence étroite

On note $\mathcal{M}_1$ l'ensemble des probabilités sur $\mathcal{B}(\mathbb{R}^d)$, C_b (resp. C_0, resp. C_k) l'ensemble des fonctions continues bornées (resp. tendant vers 0 à l'infini, resp. à support compact) sur $\mathbb{R}^d$. Soient $\mu_n, \mu \in \mathcal{M}_1$. On veut donner un sens à "μ_n converge vers μ". Il semble naturel de demander que, pour tout $A \in \mathcal{B}(\mathbb{R}^d)$, $\mu_n(A) \to \mu(A)$ mais ceci est très contraignant. Par exemple, sur $\mathbb{R}$, si $\mu_n = \delta_{\frac{1}{n}}$ et $\mu = \delta_0$, on a $\mu_n(]0,1]) = 1$ et $\mu(]0,1]) = 0$ et donc, en ce sens, μ_n ne converge pas vers μ. C'est pourquoi on introduit la notion de convergence étroite.

7.1.1. Définition.

Définition 7.1.1. *Soient $\mu_n, \mu \in \mathcal{M}_1$. On dit que μ_n converge étroitement vers μ si, pour toute $f \in C_b$, $\int f\,d\mu_n \to_n \int f\,d\mu$.*

Un critère très utile est le suivant. Rappelons que $H \subset C_0$ est total si e.v.$[H]$ est dense dans C_0 pour la norme $||f|| = \sup_x |f(x)|$.

Proposition 7.1.2. *Soient $\mu_n, \mu \in \mathcal{M}_1$. Si, pour toute $f \in H$, H total dans C_0, $\int f\,d\mu_n \to_n \int f\,d\mu$, μ_n converge étroitement vers μ.*

Preuve: Montrons d'abord que, pour toute $f \in C_0$, $\int f\,d\mu_n \to_n \int f\,d\mu$. Soit $V =$ e.v.$[H]$. On a $\overline{V} = C_0$ et, pour toute $g \in V$, $\int g\,d\mu_n \to \int g\,d\mu$. Soient $f \in C_0$ et $g \in V$, on a

$$\begin{aligned}|\int f\,d\mu_n - \int f\,d\mu| &\leq |\int f\,d\mu_n - \int g\,d\mu_n| + |\int g\,d\mu_n - \int g\,d\mu| + |\int g\,d\mu - \int f\,d\mu| \\ &\leq 2||f-g|| + |\int g\,d\mu_n - \int g\,d\mu|.\end{aligned}$$

On a donc $\limsup_n |\int f\,d\mu_n - \int f\,d\mu| \leq 2||f-g||$. Cette dernière quantité étant arbitrairement petite, $\int f\,d\mu_n \to \int f\,d\mu$.

Ceci fait, on a, pour $f \in C_b$ et $g \in C_k$, $0 \leq g \leq 1$,

$$|\int f\,d\mu_n - \int f\,d\mu| \leq |\int f\,d\mu_n - \int fg\,d\mu_n| + |\int fg\,d\mu_n - \int fg\,d\mu| + |\int fg\,d\mu - \int f\,d\mu|$$
$$\leq ||f||(1 - \int g\,d\mu_n) + |\int fg\,d\mu_n - \int fg\,d\mu| + ||f||(1 - \int g\,d\mu).$$

On a donc $\limsup_n |\int f\,d\mu_n - \int f\,d\mu| \leq 2||f||(1 - \int g\,d\mu)$. Vu qu'il existe $g_n \in C_k$, $0 \leq g_n \leq 1$, tels que $g_n \uparrow 1$ et qu'alors $\int g_n\,d\mu \uparrow \int 1\,d\mu = 1$, $1 - \int g\,d\mu$ est arbitrairement petit et $\int f\,d\mu_n \to_n \int f\,d\mu$. Ceci montre que μ_n converge étroitement vers μ. $\diamond$

Il y a deus exemples particulièrement intéressants d'ensemble total dans C_0 à savoir l'espace C_k^∞ (cor. 3.5.6) et la famille $(g_\sigma(x-a),\ \sigma > 0, a \in \mathbb{R}^d)$ (lem. 5.1.3).

7.1.2. L'exemple introductif montre que μ_n peut converger étroitement vers μ sans que $\mu_n(A)$ converge vers $\mu(A)$. La question est de savoir pour quels ensembles on a cette convergence. On note $\partial A = \overline{A} \setminus \overset{\circ}{A}$ la frontière topologique de A i.e. la fermeture moins l'intérieur.

Proposition 7.1.3. *Soient $\mu_n, \mu \in \mathcal{M}_1$. On suppose que μ_n converge étroitement vers μ. Alors, pour tout $A \in \mathcal{B}(\mathbb{R}^d)$ tel que $\mu(\partial A) = 0$, $\mu_n(A) \to \mu(A)$.*

Preuve: Il existe $f_p, g_p \in C_b^+$ telles que $g_p \downarrow 1_{\overline{A}}$, $f_p \uparrow 1_{\overset{\circ}{A}}$, alors $\int g_p\,d\mu \downarrow \mu(\overline{A})$ et $\int f_p\,d\mu \uparrow \mu(\overset{\circ}{A})$. D'où, vu l'hypothèse, $\int (g_p - f_p)\,d\mu \to_p 0$.

Soit $\varepsilon > 0$. Il existe donc $f, g \in C_b$ telles que $f \leq 1_A \leq g$ et $\int (g-f)\,d\mu < \varepsilon$. On a alors

$$\int f\,d\mu_n - \int g\,d\mu \leq \mu_n(A) - \mu(A) \leq \int g\,d\mu_n - \int f\,d\mu$$

d'où $\limsup_n |\mu_n(A) - \mu(A)| \leq \int (g-f)\,d\mu < \varepsilon$. Ceci montre que $\mu_n(A) \to \mu(A)$. $\diamond$

7.1.3. On a enfin le résultat très important suivant:

Théorème 7.1.4. *Soient $\mu_n, \mu \in \mathcal{M}_1$. La suite μ_n converge étroitement vers μ ssi, pour tout $t \in \mathbb{R}^d$, $\hat{\mu}_n(t) \to_n \hat{\mu}(t)$.*

Preuve: La condition est évidemment nécessaire puisque $f_x(t) = \mathrm{e}^{i<t,x>} \in C_b$. Réciproquement, d'après (5.5) et le théorème de Lebesgue,

$$\int g_\sigma(x-a)\,d\mu_n = (2\pi)^{-d/2} \int g_1(\sigma t)\mathrm{e}^{-i<a,t>}\hat{\mu}_n(t)\,dt$$
$$\to_n (2\pi)^{-d/2} \int g_1(\sigma t)\mathrm{e}^{-i<a,t>}\hat{\mu}(t)\,dt = \int g_\sigma(x-a)\,d\mu.$$

Puisque $H = (g_\sigma(x-a),\ \sigma > 0, a \in \mathbb{R}^d)$ est totale dans C_0, on conclut grâce à la prop. 7.1.2. $\diamond$

7.2. Convergence en loi

Dans cette section, X_n, X désignent des v.a. à valeurs $\mathbb{R}^d$. Rappelons qu'on note μ_X la loi de X et ϕ_X sa fonction caractéristique.

7.2.1. Convergence en loi des v.a..

Définition 7.2.1. *On dit qu'une suite de v.a.* X_n *converge en loi vers une probabilité* μ *(resp. une v.a.* X*) si la suite* μ_{X_n} *converge étroitement vers* μ *(resp. vers* μ_X*).*

La distinction entre convergence en loi vers μ ou vers X est une simple affaire de langage car en fait c'est la loi de X_n qui converge vers μ et donc vers la loi de X pour toute v.a. X de loi μ. Vu la prop. 7.1.2 et le th. 7.1.4, on a:

Proposition 7.2.2. *Soient* X_n *des v.a. à valeurs* $\mathbb{R}^d$ *et* $\mu \in \mathcal{M}_1$. *Il y a équivalence entre:*
(i) X_n *converge en loi vers* μ,
(ii) pour toute $f \in H$, H *total dans* C_0, $\mathbb{E}(f(X_n)) \to_n \int f\, d\mu$,
(iii) pour tout $t \in \mathbb{R}^d$, $\phi_{X_n}(t) \to_n \hat{\mu}(t)$.

En particulier X_n converge en loi vers X ssi:

$$\text{pour tout } t \in \mathbb{R}^d, \;\; \phi_{X_n}(t) = \mathbb{E}(e^{i<t,X_n>}) \to_n \phi_X(t) = \mathbb{E}(e^{i<t,X>}).$$

Proposition 7.2.3. *Si* X_n *converge en loi vers* X *et si* $\phi : \mathbb{R}^d \to \mathbb{R}^p$ *est continue,* $Y_n = \phi(X_n)$ *converge en loi vers* $Y = \phi(X)$.

Preuve: Soit $f \in C_b(\mathbb{R}^p)$, alors $f \circ \phi \in C_b(\mathbb{R}^d)$ et

$$\mathbb{E}(f(Y_n)) = \mathbb{E}(f(\phi(X_n))) \to_n \mathbb{E}(f(\phi(X))) = \mathbb{E}(f(Y)). \;\diamond$$

Enfin la prop. 7.1.3 devient:

Proposition 7.2.4. *Soit* X_n *une suite de v.a. convergeant en loi vers* μ. *Pour tout* $A \in \mathcal{B}(\mathbb{R}^d)$ *tel que* $\mu(\partial A) = 0$, *on a* $\mathbb{P}(X_n \in A) \to_n \mu(A)$.

7.2.2. Examinons le lien entre la convergence en loi et les convergences des v.a. étudiées dans la section précédente.

Proposition 7.2.5. *Si* X_n *converge en probabilité vers* X, *alors* X_n *converge en loi vers* X.

Preuve: Il suffit (prop. 7.2.2) de montrer que, pour toute $f \in C_k$, $\mathbb{E}(f(X_n)) \to_n \mathbb{E}(f(X)) = \int f d\mu_X$. Soient donc $f \in C_k$ et $\varepsilon > 0$. Il existe, f étant uniformément continue, $\alpha > 0$ tel que $|f(x) - f(y)| \leq \varepsilon$ si $|x - y| \leq \alpha$. On a alors

$$|\mathbb{E}(f(X_n)) - \mathbb{E}(f(X))| \leq \mathbb{E}(|f(X_n)) - f(X)|1_{\{|X_n - X| \leq \alpha\}})$$
$$+\mathbb{E}(|f(X_n)) - f(X)|1_{\{|X_n - X| > \alpha\}}) \leq \varepsilon + 2||f||\, \mathbb{P}(|X_n - X| > \alpha)$$

d'où $\limsup_n |\mathbb{E}(f(X_n)) - \mathbb{E}(f(X))| \le \varepsilon$ et $\mathbb{E}(f(X_n)) \to_n \mathbb{E}(f(X))$. $\diamond$

Exemple. Soir X_n une suite de v.a.r. telle que $\mathbb{P}(X_n = 1) = p_n$ et $\mathbb{P}(X_n = 0) = 1-p_n$ avec $0 < p_n < 1$. $X_n \to_n 0$ en probabilité ssi $p_n \to_n 0$, $X_n \to_n 1$ en probabilité ssi $p_n \to_n 0$ et sinon ne converge pas en probabilité tandis que, vu que $\mathbb{E}(f(X_n)) = p_n f(1) + (1-p_n)f(0)$, X_n converge en loi ssi $p_n \to_n p$. Ceci montre qu'en général la convergence en loi n'implique pas la convergence en probabilité. On a cependant:

Proposition 7.2.6. *Si X_n converge en loi vers $a \in \mathbb{R}^d$, alors X_n converge en probabilité vers a.*

Preuve: Soit $\varepsilon > 0$. On choisit $f \in C_b$ telle que $f(a) = 0$ et $f(x) = 1$ si $|x - a| \ge \varepsilon$. Alors
$$\mathbb{P}(|X_n - a| > \varepsilon) = \mathbb{E}(1_{\{|X_n - a| > \varepsilon\}}) \le \mathbb{E}(f(X_n)) \to_n f(a) = 0. \ \diamond$$
Le résultat suivant sera utile.

Proposition 7.2.7. *Soient X_n et Y_n des v.a. réelles. On suppose que X_n converge en loi vers X et que Y_n converge en loi vers $a \in \mathbb{R}$, alors (X_n, Y_n) converge en loi vers (X, a). En particulier $X_n + Y_n$ converge en loi vers $X + a$ et $X_n Y_n$ converge en loi vers aX.*

Preuve: Posons, pour $u, v \in \mathbb{R}$, $\rho_n = \mathbb{E}(\mathrm{e}^{i(uX_n + vY_n)}) - \mathbb{E}(\mathrm{e}^{i(uX + va)})$. Il suffit (prop.7.2.2) de montrer que $\rho_n \to_n 0$. On a
$$\begin{aligned} |\rho_n| &\le |\mathbb{E}[\mathrm{e}^{iuX_n}(\mathrm{e}^{ivY_n} - \mathrm{e}^{iva})]\,| + |\mathbb{E}[\mathrm{e}^{iva}(\mathrm{e}^{iuX_n} - \mathrm{e}^{iuX})]\,| \\ &\le \mathbb{E}(|\mathrm{e}^{ivY_n} - \mathrm{e}^{iva}|) + |\mathbb{E}(\mathrm{e}^{iuX_n} - \mathrm{e}^{iuX})| = a_n + b_n. \end{aligned}$$

D'une part, posant $f(y) = |\mathrm{e}^{ivy} - \mathrm{e}^{iva})|$, $f \in C_b$ et donc $a_n = \mathbb{E}(f(Y_n)) \to_n f(a) = 0$; d'autre part, par hypothèse, $b_n \to_n 0$. La fin de la proposition résulte de la prop.7.2.3. $\diamond$

7.2.3. Le cas des v.a. entières.

Proposition 7.2.8. *Soit X_n, X des v.a. à valeurs $\mathbb{N}$. Alors X_n converge en loi vers X ssi, pour tout $k \in \mathbb{N}$, $\mathbb{P}(X_n = k) \to_n \mathbb{P}(X = k)$.*

Preuve: (i) Supposons que X_n converge en loi vers X et soit $f \in C_k$ telle que $f(k) = 1$, $f = 0$ sur $]k-1, k+1[^c$. On a
$$\mathbb{P}(X_n = k) = \mathbb{E}(f(X_n)) \to_n \mathbb{E}(f(X)) = \mathbb{P}(X = k).$$

(ii) Supposons que, pour tout $k \in \mathbb{N}$, $\mathbb{P}(X_n = k) \to_n \mathbb{P}(X = k)$. On a, pour $f \in C_k$ et donc nulle hors de $]-m, +m[$,
$$\mathbb{E}(f(X_n)) = \sum_{k=-m}^{m} f(k)\mathbb{P}(X_n = k) \to_n \sum_{k=-m}^{m} f(k)\mathbb{P}(X = k) = \mathbb{E}(f(X)) = \int f\, d\mu_X.$$

On applique la prop. 7.2.2. $\diamond$

7.2.4. Convergence en loi et convergence des espérances. Soit X_n une suite de v.a. réelles intégrables convergeant en loi vers X. A-t-on $\mathbb{E}(X_n) \to_n \mathbb{E}(X)$? En général non puisque la fonction $f(x) = x$ est continue mais non bornée. Dans le sens positif, on a:

Proposition 7.2.9. *Soit X_n une suite de v.a. réelles convergeant en loi vers X. On suppose qu'il existe $\alpha > 0$ tel que $\sup_n \mathbb{E}(|X_n|^{1+\alpha}) = M < +\infty$. Alors $X \in L^1$ et $\mathbb{E}(X_n) \to_n \mathbb{E}(X)$.*

Preuve: Soit $a > 0$. On pose $f_a(x) = |x| \wedge a$, $g_a(x) = -a \vee (x \wedge a)$. Noter que $f_a, g_a \in C_b$ et que

$$|g_a(x) - x| \le |x| 1_{\{|x|>a\}} \le \frac{|x|^{1+\alpha}}{a^\alpha}.$$

D'une part

$$\mathbb{E}(|X|^{1+\alpha} \wedge a) = \mathbb{E}(f_a(X)) = \lim_n \mathbb{E}(f_a(X_n)) \le \mathbb{E}(|X_n|^{1+\alpha}) \le M$$

d'où, pour $a \uparrow +\infty$, $\mathbb{E}(|X|^{1+\alpha}) = \lim \uparrow_a \mathbb{E}(|X|^{1+\alpha} \wedge a) \le M$. D'autre part

$$\begin{aligned} |\mathbb{E}(X_n) - \mathbb{E}(X)| &\le \mathbb{E}(|X_n - g_a(X_n)|) + |\mathbb{E}(g_a(X_n)) - \mathbb{E}(g_a(X))| + \mathbb{E}(|g_a(X) - X|) \\ &\le \frac{\mathbb{E}(|X_n|^{1+\alpha})}{a^\alpha} + |\mathbb{E}(g_a(X_n)) - \mathbb{E}(g_a(X))| + \frac{\mathbb{E}(|X|^{1+\alpha})}{a^\alpha} \end{aligned}$$

d'où $\limsup_n |\mathbb{E}(X_n) - \mathbb{E}(X)| \le \frac{2M}{a^\alpha}$ et le résultat cherché a étant arbitrairement grand. $\diamond$

7.2.5. Convergence en loi et fonctions de répartition

Proposition 7.2.10. *Soient X_n une suite de v.a. réelles de fonctions de répartition F_n et μ une probabilité sur $\mathbb{R}$ de fonction de répartition F. Alors X_n converge en loi vers μ ssi, pour tout t point de continuité de F, $F_n(t) \to_n F(t)$.*

Preuve: (i) Soit t un point de continuité de F. On a donc $\mu(\{t\}) = F(t) - F(t-) = 0$. Soit $A =]-\infty, t]$, $\partial A = \{t\}$ et $\mu(\partial A) = 0$ donc (prop. 7.1.3):

$$F_n(t) = \mu_{X_n}(]-\infty, t]) \to_n \mu(]-\infty, t]) = F(t).$$

(ii) Si $F_n(t) \to_n F(t)$ pour tout t point de continuité de F, on a, les points de discontinuité de F étant au plus dénombrables puisque F est croissante, $F_n \to_n F$ λ p.p. Soient μ_n la loi de X_n et $H = C_k^1$. H étant total dans C_0, pour montrer que X_n converge en loi vers μ, il suffit (prop. 7.1.2) de montrer que $\int f\, d\mu_n \to_n \int f\, d\mu$ pour toute $f \in H$. Si $f \in H$, $f(x) = \int_{-\infty}^{x} f'(t)\, dt$ et on a (Fubini et Lebesgue):

$$\int f\, d\mu_n = \int_{-\infty}^{+\infty} \int_{-\infty}^{x} f'(t)\, dt\, d\mu_n(x) = \int_{-\infty}^{+\infty} f'(t) \int_{t}^{+\infty} d\mu_n(x)\, dt$$

$$= \int_{-\infty}^{+\infty} f'(t)(1 - F_n(t))\, dt \to_n \int_{-\infty}^{+\infty} f'(t)(1 - F(t))\, dt = \int f\, d\mu. \ \diamond$$

On en déduit un cas particulier d'un résultat dû à Skorokhod.

Corollaire 7.2.11. *Soit X_n une suite de v.a.r. convergeant en loi vers X_∞. Il existe des v.a.r. (pas nécessairement définies sur le même espace de probabilité) Y_n, $1 \le n \le +\infty$, telles que, pour $1 \le n \le +\infty$, loi de Y_n = loi de X_n et $Y_n \to_n Y_\infty$ presque sûrement.*

Preuve: Soient F_n et F les fonctions de répartition de X_n et X_∞ et $C(F)$ l'ensemble des points de continuité de F. On pose $F^{-1}(u) = \inf(t,\, F(t) \ge u)$. Soient $A = \{u \in [0,1],\ \exists t_1 \ne t_2$ tels que $F(t_1) = F(t_2) = u\}$ et $B = [0,1] \setminus A$. Noter que A est dénombrable. et que, pour tout $u \in B$, $y < F^{-1}(u) \Rightarrow F(y) < u$ et $y > F^{-1}(u) \Rightarrow F(y) > u$. On en déduit que, pour tout $u \in B$, $F_n^{-1}(u) \to_n F^{-1}(u)$. En effet soient $u \in B$ et $y \in C(F)$ tels que $y > F^{-1}(u)$, on a $F(y) > u$ et aussi (th. 7.2.10), pour n assez grand, $F_n(y) > u$ et $y \ge F_n^{-1}(u)$ ce qui implique, $C(F)$ étant dense, $\limsup_n F_n^{-1}(u) \le F^{-1}(u)$. Considérant $y \in C(F)$ tel que $y < F^{-1}(u)$, on a, par un argument symétrique que $\liminf_n F_n^{-1}(u) \ge F^{-1}(u)$. D'où $\lim_n F_n^{-1}(u) = F^{-1}(u)$ si $u \in B$. On considère alors l'espace de probabilité $([0,1], \mathcal{B}([0,1]), \lambda =$ mesure de Lebesgue$)$ et soit U la v.a. $U(u) = u$. On pose $Y_n = F_n^{-1}(U)$, $Y_\infty = F^{-1}(U)$. D'après la prop. 4.3.2, Y_n et X_n ont même loi et, pour tout $u \in B$, $Y_n(u) = F_n^{-1}(u) \to_n Y_\infty(u) = F^{-1}(u)$ et, comme $\lambda(B) = 1$, $Y_n \to_n Y_\infty$ p.s. $\diamond$

7.2.6. Théorème de Levy. S'il est souvent facile de montrer que $\phi_{X_n}(t) \to_n \phi(t)$, il est plus délicat de montrer que $\phi(t)$ est une fonction caractéristique. De plus ce n'est pas toujours vrai. Donnons un exemple. Soit X_n une suite de v.a.r. de loi uniforme sur $[-n, +n]$. On a $\phi_{X_n}(0) = 1$ et, pour $t \ne 0$,

$$\phi_{X_n}(t) = \frac{1}{2n}\int_{-n}^{n} \mathrm{e}^{itx}\, dx = \frac{\sin(nt)}{nt}.$$

Donc $\phi_{X_n}(t) \to_n 1_{\{0\}}(t)$ qui n'est pas une fonction caractéristique puisque pas continue en 0. En fait, pour $f \in C_k$, il est immédiat que $\int f\, d\mu_{X_n} \to_n 0$ et μ_{X_n} converge en un sens affaiblie vers 0. La réponse à ce problème est donnée par le théorème de Lévy.

Théorème 7.2.12. *Soit X_n une suite de v.a. telle que, pour tout $t \in \mathbb{R}^d$, $\phi_{X_n}(t) \to_n \phi(t)$. Si ϕ est continue en 0, il existe une probabilité μ sur $\mathbb{R}^d$ telle que $\hat{\mu} = \phi$ et X_n converge en loi vers μ.*

Preuve: On a besoin du résultat d'analyse suivant que nous admettons. On dit qu'une suite $\mu_n \in \mathcal{M}_b$ converge faiblement s'il existe $\mu \in \mathcal{M}_b$ telle que, pour toute $f \in C_0$, $\int f\, d\mu_n \to_n \int f\, d\mu$. Alors

Théorème 7.2.13. *Soient $\mu_n \in \mathcal{M}_b$ telles que $A = \sup_n \mu_n(\mathbb{R}^d) < +\infty$, alors il existe une sous-suite μ_{n_k} convergeant faiblement.*

Ceci fait, on note μ_n la loi de X_n. Puisque $\mu_n(\mathbb{R}^d) = 1$, il existe (th.7.2.13) une sous-suite μ_{n_k} telle que μ_{n_k} converge faiblement vers $\mu \in \mathcal{M}_b$. On pose $\mu'_k = \mu_{n_k}$. D'après (5.5), on a, pour tout $a \in \mathbb{R}^d$,

$$\int g_\sigma(x-a)\, d\mu'_k(x) = (2\pi)^{-d/2}\int \mathrm{e}^{-i<a,u>} g_1(\sigma u)\hat{\mu}'_k(u)\, du.$$

Passant à la limite en k, on a (justifier),

$$\int g_\sigma(x-a)\,d\mu(x) = (2\pi)^{-d/2}\int \mathrm{e}^{-i<a,u>} g_1(\sigma u)\phi(u)\,du.$$

On a donc vu (5.5), pour tout $a \in \mathbb{R}^d$,

$$\int \mathrm{e}^{-i<a,u>} g_1(\sigma u)\hat{\mu}(u)\,du = \int \mathrm{e}^{-i<a,u>} g_1(\sigma u)\phi(u)\,du.$$

D'où (th.5.1.2) $\hat{\mu}(u)g_1(\sigma u) = \phi(u)g_1(\sigma u)$ λ p.p. et, g_1 étant > 0, $\hat{\mu}(u) = \phi(u)$ λ p.p. Soit $E = \{\hat{\mu} = \phi\}$, on a $\lambda(E^c) = 0$. Il existe donc $x_n \in E$ tel que $x_n \to 0$. On a, pour tout n, $\hat{\mu}(x_n) = \phi(x_n)$ et, les deux fonctions étant continues en 0, $\mu(\mathbb{R}^d) = \hat{\mu}(0) = \phi(0) = \lim_n \hat{\mu}_n(0) = 1$. Donc $\mu \in \mathcal{M}_1$ et (prop. 7.1.2) μ'_k converge étroitement vers μ. On en déduit que $\phi = \hat{\mu}$ et que μ_n converge étroitement vers μ. $\diamond$

7.3. Convergence vers la loi normale

7.3.1. Le théorème de la limite centrale.

Théorème 7.3.1. *Soit X_n une suite de v.a. à valeurs $\mathbb{R}^d$ indépendantes et de même loi. On suppose que $\mathbb{E}(|X_1|^2) < +\infty$ et on pose $m = \mathbb{E}(X_1)$, $K = K(X_1)$, $S_n = X_1 + \ldots, X_n$. Alors $\frac{1}{\sqrt{n}}(S_n - nm)$ converge en loi vers $N_d(0,K)$.*

Preuve: Il suffit de considérer le cas où $m = \mathbb{E}(X_1) = 0$. On pose $\phi(t) = \phi_{X_1}(t)$. Vu la prop. 5.2.5, $\frac{\partial}{\partial t_k}\phi(0) = 0$, $\frac{\partial^2}{\partial t_j t_k}\phi(0) = -K_{j,k}$. On a donc

$$\phi(t) = 1 - \frac{1}{2}t^{\mathrm{T}}Kt + |t|^2\varepsilon(t) \text{ avec } \lim_{t\to 0}|\varepsilon(t)| = 0.$$

On en déduit

$$\phi_{\frac{1}{\sqrt{n}}S_n}(t) = \phi_{S_n}(\frac{t}{\sqrt{n}}) = (\phi(\frac{t}{\sqrt{n}}))^n = (1 - \frac{1}{2n}t^{\mathrm{T}}Kt + \frac{|t|^2}{n}\varepsilon(\frac{t}{\sqrt{n}}))^n \to_n \exp(-\frac{1}{2}t^{\mathrm{T}}Kt).$$

Ceci d'après le lem. 7.3.2 ci-dessous. Donc $\frac{1}{\sqrt{n}}S_n$ converge en loi vers $N_d(0,K)$ d'après la prop. 7.2.2.

Lemme 7.3.2. *Soient $z_n, z \in \mathbb{C}$ tels que $z_n \to_n z$, alors on a $(1+\frac{z_n}{n})^n \to_n e^z$.*

Preuve: Pour $z_n \in \mathbb{R}$, le résultat est classique. Remarquant que, pour $a, b \in \mathbb{C}$, on a $|a^n - b^n| \le n|a-b|$ si $|a| \le 1$, $|b| \le 1$, on a

$$|\frac{(1+\frac{z_n}{n})^n}{(1+\frac{|z_n|}{n})^n} - \frac{\mathrm{e}^z}{\mathrm{e}^{|z|}}| = |(\frac{1+\frac{z_n}{n}}{1+\frac{|z_n|}{n}})^n - (\frac{\mathrm{e}^{\frac{z}{n}}}{\mathrm{e}^{\frac{|z|}{n}}})^n| \le n\,|\frac{1+\frac{z_n}{n}}{1+\frac{|z_n|}{n}} - \frac{\mathrm{e}^{\frac{z}{n}}}{\mathrm{e}^{\frac{|z|}{n}}}|$$

$$\le n\,\frac{|(1+\frac{z_n}{n})\,\mathrm{e}^{\frac{|z|}{n}} - (1+\frac{|z_n|}{n})\,\mathrm{e}^{\frac{z}{n}}|}{(1+\frac{|z_n|}{n})\,\mathrm{e}^{\frac{|z|}{n}}} \le \frac{|z_n + |z| - |z_n| - z + \varepsilon(\frac{1}{n})|}{(1+\frac{|z_n|}{n})\,\mathrm{e}^{\frac{|z|}{n}}} \to_n 0.$$

Donc $\frac{(1+\frac{z_n}{n})^n}{(1+\frac{|z_n|}{n})^n} \to_n \frac{e^z}{e^{|z|}}$ et, vu que $(1+\frac{|z_n|}{n})^n \to_n e^{|z|}$, $(1+\frac{z_n}{n})^n \to_n e^z$. $\diamond$

7.3.2. Le cas réel.

Corollaire 7.3.3. *Soit X_n une suite de v.a.r. indépendantes, de même loi, de carré intégrable. On pose $S_n = X_1 + \ldots + X_n$, $m = \mathbb{E}(X_1)$, $\sigma^2 = Var(X_1)$ qu'on suppose > 0. Alors, pour $-\infty \le a < b \le +\infty$,*

$$\mathbb{P}(a < \frac{S_n - nm}{\sigma\sqrt{n}} < b) \to_n \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{t^2}{2}}\, dt.$$

Preuve: Ceci résulte du th. 7.3.1 et de la prop. 7.2.4. $\diamond$

Exemple. Soient $X_1, \ldots, X_n, \ldots$ une suite de v.a. réelles indépendantes et de même loi de Poisson $\mathcal{P}(1)$ et $S_n = X_1 + \ldots + X_n$. On sait (2.3.3) que $S_n \sim \mathcal{P}(n)$ et (2.2.5) que $\mathbb{E}(S_n) = n$, $\mathrm{Var}(S_n) = n$. Posons

$$Y_n = \frac{S_n - n}{\sqrt{n}}.$$

D'après le th. 7.3.1, Y_n converge en loi vers $Z \sim N_1(0,1)$. Soit $h(x) = (-x) \wedge 0$, h est continue donc (prop.7.2.3) $Y_n^- = h(Y_n)$ converge en loi vers $Z^- = h(Z)$. Vu que $\mathbb{E}((Y_n^-)^2) \le \mathbb{E}(Y_n^2) = \frac{1}{n}\mathrm{Var}(S_n) = 1$, on a (prop. 7.2.9) $\mathbb{E}(Y_n^-) \to_n \mathbb{E}(Z^-)$. Mais

$$\mathbb{E}(Y_n^-) = \mathbb{E}(h(Y_n)) = \sum_{k=0}^{+\infty} h(\frac{k-n}{\sqrt{n}})\mathbb{P}(S_n = k) = \sum_{k=0}^{n} \frac{n-k}{\sqrt{n}} e^{-n}\frac{n^k}{k!}$$
$$= \frac{e^{-n}}{\sqrt{n}}\{\sum_{k=0}^{n} \frac{n^{k+1}}{k!} - \sum_{k=1}^{n} \frac{n^k}{(k-1)!}\} = \frac{e^{-n}}{\sqrt{n}}\frac{n^{n+1}}{n!} = \frac{e^{-n}n^n\sqrt{n}}{n!}$$

et

$$\mathbb{E}(Z^-) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} x^- e^{-\frac{x^2}{2}}\, dx = \frac{1}{\sqrt{2\pi}}\int_0^{+\infty} x e^{-\frac{x^2}{2}}\, dx = \frac{1}{\sqrt{2\pi}}\int_0^{+\infty} d(-e^{-\frac{x^2}{2}}) = \frac{1}{\sqrt{2\pi}}$$

d'où $\frac{e^{-n}n^n\sqrt{n}}{n!} \to_n \frac{1}{\sqrt{2\pi}}$ i.e. $n! \sim \sqrt{2\pi n}\, e^{-n}n^n$ (formule de Stirling).

7.3.3. Vitesse de convergence. Pour $d = 1$, le théorème de la limite centrale nous dit que, pour n assez grand, la loi de $\frac{S_n - nm}{\sigma\sqrt{n}}$ i.e. de S_n centrée réduite est proche de la loi $N_1(0,1)$. Pour être vraiment utile, un tel résultat doit être accompagné de précisions sur la vitesse de convergence. A ce sujet, on a le théorème de Berry-Esseen que nous montrerons section 7.4.

Théorème 7.3.4. *Soit X_n une suite de v.a. indépendantes et de même loi avec $\mathbb{E}(|X_1|^3) < +\infty$. On pose $m = \mathbb{E}(X_1)$, $\sigma^2 = \mathbb{E}(X_1 - m)^2$, $\rho = \mathbb{E}(|X_1 - m|^3)$. Alors:*

$$\sup_x |\,\mathbb{P}(\frac{S_n - nm}{\sigma\sqrt{n}} \le x) - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt\,| \le \frac{\rho}{\sigma^3\sqrt{n}}.$$

Exemple. Soit $Z_n \sim B(n,p)$. On a $Z_n = \sum_{k=1}^n X_k$ avec X_k v.a. indépendantes de loi $B(1,p)$. On a, posant $q = 1-p$, $\sigma^2(X_1) = pq$, $\rho = pq(p^2+q^2) \leq pq$ et finalement

$$|\,\mathbb{P}(\frac{Z_n - np}{\sqrt{npq}} \leq x) - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^x \mathrm{e}^{-\frac{t^2}{2}}\,dt\,| \leq \frac{1}{\sqrt{pqn}}.$$

On voit que cette approximation est peu fiable pour p proche de 0 ou 1.

7.4. Complément : démonstration du théorème de Berry-Esseen.

Il s'agit de montrer le th. 7.3.4. En fait nous montrons un énoncé un peu différent où la conctante C n'est pas précisée. Cette valeur de C n'est pas connue, on sait seulement que $C \leq 0,8$.

Théorème. Il existe une constante universelle C telle que, pour toute suite X_n de v.a.r. indépendantes et de même loi avec $\mathbb{E}(|X_1|^3) < +\infty$, on ait, posant $m = \mathbb{E}(X_1)$, $\sigma^2 = \mathbb{E}(X_1 - m)^2$, $\rho = \mathbb{E}(|X_1 - m|^3)$,

$$\sup_x |\,\mathbb{P}(\frac{S_n - nm}{\sigma\sqrt{n}} \leq x) - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^x \mathrm{e}^{-\frac{t^2}{2}}\,dt\,| \leq C\,\frac{\rho}{\sigma^3\sqrt{n}}.$$

7.4.1. Preuve: (D'après Ho et Chen reprenant une méthode de Stein).

On fixe n et on pose $Y_i = \frac{X_i - m}{\sigma\sqrt{n}}$, $U_n = \sum_1^n Y_i$ μ = loi de Y_i. On a $\mathbb{E}(Y_i) = 0$, $n\,\mathbb{E}(Y_i^2) = 1$, $n^{3/2}\mathbb{E}(|Y_i|^3) = \rho$, $\sqrt{n}\,\mathbb{E}(|Y_1| \leq ||\sqrt{n}Y_1||_3 \leq ||\sqrt{n}Y_1||_3^3 = \rho$ puisque $||\sqrt{n}Y_1||_3 \geq ||\sqrt{n}Y_1||_2 = 1$. On note

$$\phi(x) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}, \quad \Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^x \mathrm{e}^{-\frac{t^2}{2}}\,dt.$$

Il s'agit de montrer que

$$\sup_x |\mathbb{P}(U_n \leq x) - \Phi(x)| \leq C\frac{\rho}{\sqrt{n}}. \tag{7.1}$$

On considère, pour $b \in \mathbb{R}$, notant $N(h) = \int h(t)\phi(t)\,dt$,

$$f_b(x) = \mathrm{e}^{\frac{x^2}{2}}\int_{-\infty}^x \mathrm{e}^{-\frac{t^2}{2}}(h_b(t) - N(h_b))\,dt, \quad h_b = 1_{]-\infty,b]}\,. \tag{7.2}$$

La fonction f_b est dérivable en tout $x \neq b$, d'où, posant $f_b'(b) = bf(b) + 1 - N(h_b)$,

$$\text{pour tout } x \in \mathbb{R}, \quad f_b'(x) - xf_b(x) = h_b(x) - N(h_b). \tag{7.3}$$

On a donc

$$\mathbb{P}(U_n \leq b) - \Phi(b) = \mathbb{E}(f_b'(U_n) - U_n f_b(U_n)). \tag{7.4}$$

7.4.2. On admet pour l'instant que

$$\text{pour tout } x \in \mathbb{R}, \quad |f_b(x)| \leq 1, \quad |f_b'(x)| \leq 1. \tag{7.5}$$

On dira que $f \in \mathcal{C}$ si $f \in \mathcal{B}(\mathbb{R})$ et, s'il existe $f' \in \mathcal{B}(\mathbb{R})$ telle que, pour tous $x < y$, $f(y) - f(x) = \int_x^y f'(t)\,dt$. Soit $f \in \mathcal{C}$. Vu la symétrie, l'indépendance et Fubini,

$$\begin{aligned}\mathbb{E}(U_n f(U_n)) &= \sum_{i=1}^n \mathbb{E}(Y_i\, f(\sum_{j\neq i} Y_j + Y_i)) = n\,\mathbb{E}(Y_n f(\sum_{i=1}^{n-1} Y_i + Y_n))\\ &= n\int \mathbb{E}(sf(U_{n-1}+s))\,d\mu(s)) = n\int \mathbb{E}(s(f(U_{n-1}+s) - f(U_{n-1}))\,d\mu(s)\\ &= n\mathbb{E}(\int_{s\geq 0}\int_0^s f'(U_{n-1}+t)\,dt\,d\mu(s) - n\mathbb{E}(\int_{s<0}\int_s^0 f'(U_{n-1}+t)\,dt\,d\mu(s).\end{aligned}$$

On obtient donc, posant

$$K(t) = n\int_{[t,+\infty[} s\,d\mu(s),\ t \geq 0, \quad K(t) = -n\int_{]-\infty,t]} s\,d\mu(s),\ t < 0, \tag{7.6}$$

$$\mathbb{E}(U_n f(U_n)) = \mathbb{E}(\int f'(U_{n-1}+t)K(t)\,dt), \quad f \in \mathcal{C}. \tag{7.7}$$

Vu l'indépendance, (7.4) peut s'écrire:

$$\mathbb{P}(U_n \leq b) - \Phi(b) = \mathbb{E}(\int\int [\, f_b'(U_{n-1}+s) - f_b'(U_{n-1}+t)\,]\,K(t)\,dt\,d\mu(s)). \tag{7.8}$$

Donnons quelques propriétés de $K(t)$.

Lemme 7.4.1. *$K(t)$ est une densité de probabilité vérifiant $\int |t|K(t)\,dt = \frac{\rho}{2\sqrt{n}}$ et $\int_{\{|t|\leq\rho/\sqrt{n}\}} K(t)\,dt \geq 1/2$.*

Preuve: Evidemment $K(t) \geq 0$. Par Fubini sur $\mathbb{R}^+$ et $\mathbb{R}^-$, $\int |t|^r K(t)\,dt = \frac{n}{r+1}\int |s|^{r+2}\,d\mu(s)$. D'où $\int K(t)\,dt = n\mathbb{E}(Y_1^2) = 1$ et $\int |t|K(t)\,dt = \frac{n}{2}\mathbb{E}(|Y_1|^3) = \frac{\rho}{2\sqrt{n}}$. Enfin

$$\int_{\{|t|>\rho/\sqrt{n}\}} K(t)\,dt \leq \frac{\sqrt{n}}{\rho}\int_{\{|t|>\rho/\sqrt{n}\}} |t|K(t)\,dt \leq \frac{\sqrt{n}}{\rho}\int |t|K(t)\,dt = \frac{1}{2}. \ \diamond$$

La preuve repose sur une inégalité de concentration pour U_{n-1}.

Lemme 7.4.2. *On a, pour tous $a < b$, $\mathbb{P}(a < U_{n-1} < b) \leq b - a + 2\rho/\sqrt{n}$.*

Preuve: On considère la fonction f définie par $f(x) = -\frac{b-a}{2} - \frac{\rho}{\sqrt{n}}$ si $x \leq a - \frac{\rho}{\sqrt{n}}$, $f(x) = x - \frac{b+a}{2}$ si $a - \frac{\rho}{\sqrt{n}} \leq x \leq b + \frac{\rho}{\sqrt{n}}$ et $f(x) = \frac{b-a}{2} + \frac{\rho}{\sqrt{n}}$ si $x \geq b + \frac{\rho}{\sqrt{n}}$. On a

$|f(x)| \leq \frac{b-a}{2} + \frac{\rho}{\sqrt{n}}$ et $f \in \mathcal{C}$ avec $f'(x) = 1_{\{a-\frac{\rho}{\sqrt{n}} \leq x \leq b+\frac{\rho}{\sqrt{n}}\}}$. On a alors, vu (7.7), le lem. 7.4.1 et que $\mathbb{E}(|U_n|) \leq \{\mathbb{E}(U_n^2)\}^{1/2} = 1$,

$$\mathbb{P}(a < U_{n-1} < b) \leq 2\mathbb{E}(\int_{\{|t| \leq \rho/\sqrt{n}\}} 1_{\{a<U_{n-1}<b\}} K(t)\, dt\,)$$

$$\leq 2\mathbb{E}(\int 1_{\{a-\frac{\rho}{\sqrt{n}}<U_{n-1}+t<b+\frac{\rho}{\sqrt{n}}\}} K(t)\, dt\,) = 2\mathbb{E}(\int f'(U_{n-1}+t)K(t)\, dt)$$

$$= 2\mathbb{E}(U_n f(U_n)) \leq 2||f||_\infty\, ||U_n||_1 \leq b - a + 2\frac{\rho}{\sqrt{n}}. \ \diamond$$

On peut maintenant exploiter (7.8). Remarquons d'abord que, vu (7.5),

$$|f_b'(u+s) - f_b'(u+t)| \leq |(u+s)f_b(u+s) - (u+t)f_b(u+t)| + |h_b(u+s) - h_b(u+t)|$$
$$\leq |u|\, |f_b(u+s) - f_b(u+t)| + |sf_b(u+s)| + |tf_b(u+t)| + |h_b(u+s) - h_b(u+t)|$$
$$\leq (|u|+1)(|t|+|s|) + 1_{\{s \geq t\}} 1_{\{b-s \leq u \leq b-t\}} + 1_{\{s<t\}} 1_{\{b-t \leq u \leq b-s\}}.$$

Reportant ceci dans (7.8), on obtient, utilisant le lem. 7.4.2, que $\int |t| K(t)\, dt = \frac{\rho}{2\sqrt{n}}$, que $\int |s|\, d\mu(s) = \mathbb{E}(|Y_1|) \leq \frac{\rho}{\sqrt{n}}$ et que $\mathbb{E}(|U_{n-1}|) \leq \{\mathbb{E}(U_{n-1}^2)\}^{1/2} \leq 1$,

$$\sup_b |\mathbb{P}(U_n \leq b) - \Phi(b)| \leq \int\int (|s|+|t|)(\mathbb{E}(|U_{n-1}|+1)K(t)\, d\mu(s)\, dt$$
$$+ \int\int 1_{\{s \geq t\}} \mathbb{P}(b-s \leq U_{n-1} \leq b-t) d\mu(s)\, dt + \int\int 1_{\{s<t\}} \mathbb{P}(b-t \leq U_{n-1} \leq b-s) d\mu(s)\, dt$$
$$\leq 3\int\int (|s|+|t|)K(t)\, d\mu(s)\, dt + \frac{2\rho}{\sqrt{n}} \leq \frac{13}{2}\frac{\rho}{\sqrt{n}}.$$

7.4.3. Il reste à montrer (7.5). On a les inégalités classiques suivantes:

$$\text{pour } x \geq 0,\ \phi(x) \geq x(1-\Phi(x)), \qquad \text{pour } x \leq 0,\ \phi(x) \geq |x|\, \Phi(x).$$

En effet, pour $x > 0$, on a (dériver) :

$$\frac{\phi(x)}{x} = \frac{1}{\sqrt{2\pi}} \int_x^{+\infty} (1 + \frac{1}{t^2}) \mathrm{e}^{-t^2/2}\, dt \geq 1 - \Phi(x).$$

Par symétrie on obtient le cas $x < 0$.

On suppose $b \geq 0$. Le cas $b < 0$ se traite de façon analogue mais on voit facilement, remplaçant U_n par $-U_n$, qu'il suffit de montrer (7.1) pour $x \geq 0$. On remarque d'abors que:

$$\text{pour } x \geq b,\ f_b(x) = \frac{\Phi(b)(1-\Phi(x))}{\phi(x)}; \quad \text{pour } x \leq b,\ f_b(x) = \frac{\Phi(x)(1-\Phi(b))}{\phi(x)}.$$

(i) On suppose $x > b$. Alors $f_b'(x) = \Phi(b)(\frac{x(1-\Phi(x))}{\phi(x)} - 1)$ d'où $-1 \leq f_b'(x) \leq 0$.

(ii) On suppose $0 \leq x < b$. Alors $f_b'(x) = 1 - \Phi(b) + \frac{x\Phi(x)}{\phi(x)}(1 - \Phi(b))$ d'où $0 \leq f_b'(x) \leq 1 - \Phi(b) + \frac{x\Phi(x)}{\phi(x)}(1 - \Phi(x)) \leq 1 - \Phi(b) + \Phi(x) \leq 1$.

(iii) On suppose $x < 0 \leq b$. Alors $f_b'(x) = (1 - \Phi(b))(1 + \frac{x\Phi(x)}{\phi(x)})$ d'où $0 \leq f_b'(x) \leq 1 - \Phi(b) \leq 1$.

Le calcul précedent montre que $f_b(x)$ atteint son maximum en b. On a donc $0 \leq f_b(x) \leq \frac{\Phi(b)(1-\Phi(b))}{\phi(b)} \leq 1$. En effet $\frac{\Phi(b)(1-\Phi(b))}{\phi(b)} \leq \frac{\Phi(b)}{b} \leq 1$ si $b \geq b_0$ avec $b_0 \leq 0,8$ et, pour $0 \leq b \leq b_0$, $\frac{\Phi(b)(1-\Phi(b))}{\phi(b)} \leq \frac{1}{4\phi(b)} \leq \frac{1}{4\phi(b_0)} \leq \frac{1}{4\phi(0,8)} \leq 1$. $\diamond$

7.5. Complément: comportement asymptotique de la médiane empirique.

La lecture de cette section suppose que l'on a lu la section 4.9. Soit μ une probabilité sur $\mathbb{R}$. On note F sa fonction de répartition (def. 4.3.1). On sait que F est continue ssi $\mu(\{x\}) = 0$ pour tout $x \in \mathbb{R}$.

7.5.1. Médiane. Tout réel λ tel que $\mu(]-\infty, \lambda]) \geq \frac{1}{2}$ et $\mu([\lambda, +\infty[) \geq \frac{1}{2}$ s'appelle la médiane de μ. On a donc, X étant une v.a. de loi μ,

$$\mathbb{P}(X \leq \lambda) \geq \frac{1}{2} \text{ et } \mathbb{P}(X \geq \lambda) \geq \frac{1}{2}$$

i.e. $F(\lambda) \geq \frac{1}{2}$ et $F(\lambda-) \leq \frac{1}{2}$. Il y a donc trois cas possibles.

(i) Il existe un unique λ tel que $F(\lambda) = \frac{1}{2}$. Ce nombre λ est alors l'unique médiane. En particulier, c'est le cas si F est continue strictement croissante.

(ii) Il existe une infinité de λ tel que $F(\lambda) = \frac{1}{2}$. Tous ces nombres λ sont des médianes et ce sont les seuls.

(iii) Il existe λ (évidemment unique) tel que $F(\lambda-) \leq \frac{1}{2}$ et $F(\lambda) > \frac{1}{2}$. Ce nombre λ est l'unique médiane.

7.5.2. On considère maintenant une suite $X_1, \ldots, X_n, \ldots$ de v.a.r. indépendantes de même loi μ. On suppose que F fonction de répartition de μ est continue. Soit M_n la médiane empirique de $X_1, \ldots, X_{2n+1}$ (voir (4.29)).

Proposition 7.5.1. *On suppose qu'il existe un unique λ tel que $F(\lambda) = \frac{1}{2}$. Alors $M_n \to_n \lambda$ p.s.*

Preuve: Soient $s < \lambda < t$ et $F_n(u) = \frac{1}{n} \sum_{i=1}^n 1_{]-\infty,u]}(X_i)$. Noter que p.s. $F_{2n+1}(M_n) = \frac{n+1}{2n+1}$ et que (vu l'unicité de λ) $F(s) < F(\lambda) < F(t)$. Vu le th.6.4.1, $F_{2n+1}(s) \to_n F(s) < \frac{1}{2}$ et $F_{2n+1}(t) \to_n F(t) > \frac{1}{2}$ p.s. et donc $1_{]s,t]}(M_n) \to_n 1$ p.s. On en déduit que p.s. $\liminf_n M_n \geq \lambda$ et $\limsup_n M_n \leq \lambda$ i.e. $M_n \to_n \lambda$ p.s. $\diamond$

Théorème 7.5.2. *On suppose que μ a une densité $p(x)$, qu'il existe un unique λ tel que $F(\lambda) = \frac{1}{2}$, que p est continue en λ et que $p(\lambda) > 0$. Alors $Z_n = \sqrt{2n+1}(M_n - \lambda)$ converge en loi vers $N_1(0, \frac{1}{4p^2(\lambda)})$.*

Preuve: Nous allons montrer que la densité $g_n(u)$ de Z_n converge vers celle de $N_1(0, \frac{1}{4p^2(\lambda)})$ uniformément sur tout compact, ce qui montrera le théorème vu la prop 7.2.2 en choisissant $H = C_k$. D'après (4.33), la densité de M_n est:

$$\frac{(2n+1)!}{(n!)^2}(F(t))^n(1-F(t))^n p(t).$$

Un changement de variable montre que celle de Z_n est:

$$g_n(u) = \alpha_n \, . \, \{\psi_n(u)\}^n \, . \, p(\lambda + \frac{u}{\sqrt{2n+1}})$$

$$\alpha_n = \frac{(2n+1)!}{(n!)^2\sqrt{2n+1}}\frac{1}{4^n}, \quad \psi_n(u) = 4F(\lambda + \frac{u}{\sqrt{2n+1}})(1 - F(\lambda + \frac{u}{\sqrt{2n+1}})).$$

Utilisant la formule de Stirling $n! \sim (\frac{n}{e})^n \sqrt{2\pi n}$, on voit que $\alpha_n \to_n \sqrt{\frac{2}{\pi}}$. Fixons $A > 0$. L'écriture $\phi_n(u) = o(\frac{1}{a_n})$ signifie que $a_n \, \phi_n(u) \to_n 0$ uniformément en $|u| \le A$. On a alors, puisque $F' = p$ et $F(\lambda) = \frac{1}{2}$,

$$\begin{aligned} 2F(\lambda + \frac{u}{\sqrt{2n+1}}) &= 1 + \frac{u}{\sqrt{2n+1}}\, p(\lambda)\,(1 + o(1)) \\ 2(1 - F(\lambda + \frac{u}{\sqrt{2n+1}})) &= 1 - \frac{u}{\sqrt{2n+1}}\, p(\lambda)\,(1 + o(1)), \end{aligned}$$

d'où

$$n \log \psi_n(u) = n(-\frac{4u^2}{2n+1}\, p^2(\lambda) + o(\frac{1}{n})) = -2u^2p^2(\lambda) + o(1).$$

Finalement

$$g_n(u) \to_n \frac{2p(\lambda)}{\sqrt{2\pi}}\, \mathrm{e}^{-2u^2p^2(\lambda)} \ \text{ uniformément en } \ |u| \le A.$$

Mais cette dernière expression est la densité de $N_1(0, \sigma^2)$ pour $\sigma^2 = \frac{1}{4p^2(\lambda)}$. $\diamond$

7.5.3. Dans bien des cas, le th. 7.5.2 peut remplacer avantageusement le th. 7.3.1. Par exemple soit $X_1, \ldots, X_{2n+1}$ un $2n+1$ échantillon de la loi de Cauchy de densité

$$p_\theta(x) = \frac{1}{\pi(1 + (x-\theta)^2)}.$$

Cette loi n'a pas de moyenne mais a θ pour médiane. De plus $p_\theta(\theta) = \frac{1}{\pi}$. Dans ce cas $M_n \to_n \theta$ p.s. et $\sqrt{2n+1}(M_n - \theta)$ tend en loi vers $N_1(0, \frac{\pi^2}{4})$.

Plus généralement soit $p(x)$ une fonction définie sur $\mathbb{R}$, positive, paire, continue au voisinage de 0 et d'intégrale 1. On suppose que $a = p(0) > 0$ et que $\int x^2 p(x)\, dx = \sigma^2 <$

$+\infty$. On considère un $2n+1$ échantillon de la loi de densité $p_\theta(x) = p(x-\theta)$. cette loi a pour moyenne θ et pour médiane θ. Pour estimer θ, on peut utliser aussi bien $\overline{X}_{2n+1} = \frac{1}{2n+1}\sum_{i=1}^{2n+1} X_i$ que M_n. Pour comparer ces estimateurs, on peut observer que, d'après les th. 7.3.1 et 7.5.2, $\overline{X}_{2n+1}$ et M_n sont, pour n assez grand, approximativement gaussiens de moyenne θ et de variances $\frac{\sigma^2}{2n+1}$ et $\frac{1}{4a^2(2n+1)}$. On peut, suivant les cas, préférer l'un ou l'autre.

Chapitre 8

Notions de statistique

8.1. Echantillon. Modèle statistique

8.1.1. Répartition empirique. Soit μ une probabilité sur $\mathbb{R}^d$.

Définition 8.1.1. *On appelle échantillon de taille n (ou n-échantillon) de la loi μ une suite $X_1, \ldots, X_n$ de n v.a. indépendantes et de loi μ.*

On appelle réalisation du n-échantillon le résultat de n tirages indépendants selon la loi μ. C'est une suite $x_1, \ldots, x_n$ de $\mathbb{R}^d$.

Par extension, on appelle échantillon de taille infinie de la loi μ une suite de $(X_n,\, n \geq 1)$ de v.a. indépendantes et de loi μ.

Définition 8.1.2. *Soit $X = (X_1, \ldots, X_n, \ldots)$ un échantillon de taille infinie de la loi μ. La probabilité (aléatoire)*

$$\mu_n^X = \frac{1}{n} \sum_{k=1}^{n} \delta_{X_k} \tag{8.1}$$

s'appelle la répartition empirique d'ordre n de μ.

On a alors:

Proposition 8.1.3. *Presque sûrement, μ_n^X converge étroitement vers μ.*

Preuve: D'après la loi des grands nombres, pour toute $f \in C_b$,

$$\int f\, d\mu_n^X = \frac{1}{n} \sum_{k=1}^{n} f(X_k) \to_n \mathbb{E}(f(X_1)) = \int f\, d\mu \;\text{ p.s.}$$

Soit $\Phi = \{\phi_1, \ldots, \phi_p, \ldots\}$ un ensemble dense dans C_0. On a p.s. $\int \phi_p\, d\mu_n^X \to_n \int \phi_p\, d\mu$ pour tout p et donc (prop. 7.1.2) p.s. μ_n^X converge étroitement vers μ. $\diamond$

8.1.2. Le cas réel. On suppose $d = 1$ et on note F la fonction de répartition de μ. La fonction de répartition de μ_n^X s'appelle la fonction de répartition empirique de μ et se note F_n^X. On a donc

$$F_n^X(t) = \mu_n^X(]-\infty, t\,]) = \frac{1}{n}\sum_{k=1}^{n} 1_{]-\infty,t]}(X_k). \tag{8.2}$$

Il résulte de (8.2) que $nF_n^X(t) \sim B(n, F(t))$ et que, pour tout t, $F_n^X(t) \to_n F(t)$ p.s. En fait, on a un résultat beaucoup plus fort appelé théorème de Glivenko-Cantelli:

Théorème 8.1.4. $\sup_{t\in\mathbb{R}} |F_n^X(t) - F(t)| \to_n 0$ *p.s.*

Preuve: On pose $F_n = F_n^X$.

(i) On suppose que μ est la loi uniforme sur $[0,1]$. D'après (8.2) et la loi des grands nombres, il existe $A \in \mathcal{A}$ avec $\mathbb{P}(A) = 1$ tel que, pour tout $\omega \in A$, tout $k \geq 0$ et tout $p > 0$, $F_n(\frac{k}{p}) \to_n F(\frac{k}{p})$. On a alors, pour $\omega \in A$, pour $k = 1, \ldots, p$ et pour $t \in [\frac{k-1}{p}, \frac{k}{p}]$,

$$F_n(\frac{k-1}{p}) - \frac{k-1}{p} - \frac{1}{p} = F_n(\frac{k-1}{p}) - \frac{k}{p} \leq F_n(t) - t \leq F_n(\frac{k}{p}) - \frac{k-1}{p} = F_n(\frac{k}{p}) - \frac{k}{p} + \frac{1}{p}$$

d'où

$$\sup_{0\leq t\leq 1} |F_n(t) - t| \leq \max_{1\leq k\leq p} |F_n(\frac{k}{p}) - \frac{k}{p}| + \frac{1}{p}$$

et $\limsup_n \sup_{0\leq t\leq 1} |F_n(t) - t| \leq \frac{1}{p}$. Comme p est arbitraire, ceci implique que $\sup_{0\leq t\leq 1} |F_n(t) - t| \to_n 0$.

(ii) On suppose qu'il existe des v.a. $U_1, \ldots, U_n, \ldots$ indépendantes et de loi $U(0,1)$ telles que $X_n = F^{-1}(U_n)$ où $F^{-1}(u) = \inf(t,\, F(t) \geq u)$. Rappelons (voir(4.15)) que $u \leq F(t)$ ssi $F^{-1}(u) \leq t$. On note G la fonction de répartition de $U(0,1)$ et on pose $G_n = \frac{1}{n}\sum_{k=1}^{n} 1_{]-\infty,t]}(U_k)$. Vu que $U_k \leq F(t)$ ssi $X_k \leq t$, on a

$$F_n(t) - F(t) = \frac{1}{n}\sum_{k=1}^{n} 1_{]-\infty,t]}(X_k) - F(t) = \frac{1}{n}\sum_{k=1}^{n} 1_{]-\infty,F(t)]}(U_k) - F(t) = G_n(F(t)) - F(t).$$

On a donc $\sup_{t\in\mathbb{R}} |F_n(t) - F(t)| = \sup_{t\in\mathbb{R}} |G_n(F(t)) - F(t)| \leq \sup_{0\leq t\leq 1} |G_n(t) - t|$ avec égalité si F est continue car alors $F(\mathbb{R}) \supset]0,1[$. Ceci montre que $\sup_{t\in\mathbb{R}} |F_n(t) - F(t)| \to_n 0$ p.s. et que sa loi est indépendante de F si F est continue.

(iii) En fait on ne peut pas toujours écrire que $X_n = F^{-1}(U_n)$ mais il existe un espace de probabilité $(\Omega', \mathcal{A}', \mathbb{P}')$ et, sur cet espace, des v.a. $U'_1, \ldots, U'_n, \ldots$ indépendantes et de loi $U(0,1)$ telles que les v.a. $X'_n = F^{-1}(U'_n)$ soient indépendantes et de même loi que X_n (prop. 4.3.2). On conclut alors grâce à:

Lemme 8.1.5. *Soient, pour $i = 1, 2$, $(X_n^i,\, n \geq 1)$ des v.a.r. définies sur $(\Omega^i, \mathcal{A}^i, \mathbb{P}^i)$ telles que, pour tout n, $(X_1^1, \ldots, X_n^1)$ et $(X_1^2, \ldots, X_n^2)$ aient même loi et $\Phi_n \in \mathcal{B}^+(\mathbb{R}^n)$. Alors, si $\Phi_n(X_1^1, \ldots, X_n^1) \to_n 0$ $\mathbb{P}^1$ p.s., $\Phi_n(X_1^2, \ldots, X_n^2) \to_n 0$ $\mathbb{P}^2$ p.s.*

Preuve: Ceci résulte de ce que $Z_n^i = \Phi_n(X_1^i,\dots,X_n^i) \to_n 0$ $\mathbb{P}^i$ p.s ssi, pour tout $\varepsilon > 0$,

$$\sup_m \mathbb{P}^i(\max_{n\le k\le n+m} |Z_n^i| > \varepsilon) \to_n 0. \ \diamond$$

8.1.3. Moments empiriques. Soit μ une probabilité sur $\mathbb{R}$ telle que $\int |x|^p \, d\mu < +\infty$, $p \ge 2$. On note $m = \int x \, d\mu(x)$, $\sigma^2 = \int (x-m)^2 \, d\mu(x)$. On pose, pour $r \in \mathbb{N}$, $r \le p$,

$$M_n^r = \int x^r \, d\mu_n^X(x) = \frac{1}{n}\sum_{k=1}^n X_k^r. \tag{8.3}$$

Alors M_n^r s'appelle le moment empirique d'ordre r. En particulier, on note

$$\overline{X}_n = M_n^1 = \frac{1}{n}\sum_{k=1}^n X_k, \tag{8.4}$$

quantité qui s'appelle la moyenne empirique. On a

$$\mathbb{E}(\overline{X}_n) = m, \quad \mathrm{Var}(\overline{X}_n) = \frac{1}{n^2}\sum_{k=1}^n \mathrm{Var}(X_k) = \frac{\sigma^2}{n}$$

et (loi des grands nombres) $\overline{X}_n \to_n m$ p.s.

Lemme 8.1.6. *Soient $a, x_1, \dots, x_n \in \mathbb{R}$ et $\overline{x} = \frac{1}{n}\sum_{k=1}^n x_k$. Alors*

$$\sum_{k=1}^n (x_k - \overline{x})^2 = \sum_{k=1}^n (x_k - a)^2 - n(\overline{x} - a)^2 = \sum_{k=1}^n x_k^2 - n(\overline{x})^2.$$

Preuve: Il suffit de noter que $\sum (x_k - \overline{x}) = 0$ et d'écrire $x_k - \overline{x} = x_k - a + a - \overline{x}$. $\diamond$

Soit $\hat{s}_n^2$ la variance de la répartition empirique μ_n^X. On a, vu le lem.8.1.6,

$$\hat{s}_n^2 = \frac{1}{n}\sum_{k=1}^n X_k^2 - (\overline{X}_n)^2 = \frac{1}{n}\sum_{k=1}^n (X_k - \overline{X}_n)^2 = \frac{1}{n}\sum_{k=1}^n (X_k - m)^2 - (\overline{X}_n - m)^2$$

et $\mathbb{E}(\hat{s}_n^2) = \sigma^2 - \frac{\sigma^2}{n} \ne \sigma^2$. C'est pourquoi on préfère en général appelé variance empirique la quantité

$$s_n^2 = \frac{1}{n-1}\sum_{k=1}^n (X_k - \overline{X}_n)^2 \tag{8.5}$$

qui vérifie $\mathbb{E}(s_n^2) = \sigma^2$. Noter (lem. 8.1.6) que

$$s_n^2 = \frac{1}{n-1}\sum_{k=1}^n X_k^2 - \frac{n}{n-1}(\overline{X}_n)^2 \to_n \mathbb{E}(X_1^2) - m^2 = \sigma^2 \ \text{ p.s.}$$

Si n est fixé, on écrit simplement $\overline{X}$ et s^2 pour $\overline{X}_n$ et s_n^2.

8.1.4. Modèle statistique. Soit $X = (X_1, \ldots, X_n)$ un n-échantillon d'une loi μ sur $\mathbb{R}$. En statistique, la loi μ est totalement ou partiellement inconnue, ce qu'on modélise en disant que μ appartient à la famille $(\mu_\theta,\ \theta \in \Theta)$. Dans ce polycopié, le plus souvent on aura $\Theta \subset \mathbb{R}^p$. Alors $X = (X_1, \ldots, X_n)$ est une v.a. de loi $\mu_\theta^{\otimes n}$. Ceci est un cas particulier de la situation plus générale suivante.

Définition 8.1.7. *On appelle modèle statistique un terme* $(\mathcal{X}, \mathcal{A}, (\mathbb{P}_\theta)_{\theta\in\Theta})$ *où* $(\mathbb{P}_\theta)_{\theta\in\Theta}$ *est une famille de probabilités sur l'espace mesurable* $(\mathcal{X}, \mathcal{A})$.

L'ensemble Θ s'appelle l'espace des paramètres et on note X l'application identique de $\mathcal{X}$ dans $\mathcal{X}$. On appellera statistique à valeurs $(E, \mathcal{E})$ toute application mesurable de $(\mathcal{X}, \mathcal{A})$ dans $(E, \mathcal{E})$. Evidemment, pour chaque $\theta \in \Theta$, $(\mathcal{X}, \mathcal{A}, \mathbb{P}_\theta)$ est un espace de probabilité. On note alors $\mathbb{E}_\theta$ l'espérance pour $\mathbb{P}_\theta$. Très grossièrement le problème est le suivant. On tire $x \in \mathcal{X}$ selon $\mathbb{P}_\theta$, $\theta \in \Theta$ étant inconnu et, à la vue du point x tiré, on cherche à dire quelque chose sur θ.

Exemple. Soit $X_1, \ldots, X_n)$ un n-échantillon de la loi $N_1(m, \sigma^2)$, m et σ^2 étant inconnus. Décrivons le modèle statistique correspondant. On a

$$\mathcal{X} = \mathbb{R}^n,\ \mathcal{A} = \mathcal{B}(\mathbb{R}^n),\ \theta = (m, \sigma^2),\ \Theta = \mathbb{R}\times]0, \infty[,\ \mathbb{P}_\theta = q_\theta.\lambda$$

avec

$$q_\theta(x_1, \ldots, x_n) = (2\pi\sigma^2)^{-n/2} \exp(-\frac{1}{2\sigma^2}\sum_{k=1}^{n}(x_k - m)^2).$$

Plus généralement:

Définition 8.1.8. *Soit* $(\mu_\theta,\ \theta \in \Theta)$ *une famille de probabilités sur* $\mathbb{R}^d$. *On appelle modèle statistique associé à un échantillon de taille infinie de* μ_θ *le modèle* $(\mathcal{X}, \mathcal{A}, (\mathbb{P}_\theta)_{\theta\in\Theta})$ *où*

$$\mathcal{X} = (\mathbb{R}^d)^{\mathbb{N}},\ x = (x_1,\ \ldots, x_n, \ldots),\ \ X_n(x) = x_n,\ \mathcal{A} = \sigma(X_n,\ n \geq 1)$$

et où, pour chaque $\theta \in \Theta$, $\mathbb{P}_\theta$ *est une probabilité sur* $(\mathcal{X}, \mathcal{A})$ *telle que les v.a.* $X_1,\ \ldots, X_n, \ldots$ *soient indépendantes et de loi* μ_θ.

On admet l'existence d'une telle probabilité $\mathbb{P}_\theta$ qui est unique vu le cor. 3.2.3 appliqué à $\mathcal{C} = \cup_n \sigma(X_1, \ldots, X_n)$.

8.2. Estimation

Soient $(\mathcal{X}, \mathcal{A}, (\mathbb{P}_\theta)_{\theta\in\Theta})$ un modèle statistique et f une application mesurable de Θ dans $\mathbb{R}$. On veut estimer $f(\theta)$ à la vue de $x \in \mathcal{X}$ résultat d'un tirage selon $\mathbb{P}_\theta$, θ inconnu. Un estimateur de $f(\theta)$ est donc une application mesurable T de $\mathcal{X}$ dans $\mathbb{R}$. Si on a tiré x, on estime $f(\theta)$ par $T(x)$. Il reste à préciser ce qu'est un "bon" estimateur.

8.2.1. Risque quadratique.

Définition 8.2.1. *Soit T un estimateur de $f(\theta)$. On appelle risque quadratique de T la fonction*

$$R_T(\theta) = \mathbb{E}_\theta[(T - f(\theta))^2]. \tag{8.6}$$

Soient S et T deux estimateurs de $f(\theta)$. On dit que T est au moins aussi bon que S si, pour tout $\theta \in \Theta$, $R_T(\theta) \leq R_S(\theta)$. On dit T est meilleur que S s'il est au moins aussi bon et si, pour un $\theta \in \Theta$, $R_T(\theta) < R_S(\theta)$. Enfin on dit que T est admissible s'il n'existe pas un meilleur estimateur. Il faut noter que comparer des estimateurs, c'est comparer des fonctions de θ et, qu'en général, il n'y a aucune raison pour que l'un soit meilleur que l'autre. Par exemple, soit $a \in \Theta$ et $T = f(a)$. Alors $R_T(a) = 0$ et, en a, cet estimateur aura un risque plus faible que tous les autres alors que, pour d'autres valeurs de θ, son risque sera élevé. Pour avoir un estimateur optimal, on est donc amené à restreindre la classe des estimateurs considérés. C'est pourquoi on introduit:

Définition 8.2.2. *On dit que T est un estimateur sans biais de $f(\theta)$ (en abrégé e.s.b.) si, pour tout $\theta \in \Theta$, $\mathbb{E}_\theta(T) = f(\theta)$.*

C'est une qualité qu'il est naturel d'imposer à un estimateur. Cependant cette condition est assez contraignante ce qui est un avantage (on aura assez facilement des estimateurs sans biais optimaux parmi les e.s.b.) et un inconvénient (on laisse échapper de très bons estimateurs).

Si T est un e.s.b. de $f(\theta)$, alors

$$R_T(\theta) = \mathbb{E}_\theta[(T - f(\theta))^2] = \mathbb{E}_\theta[(T - \mathbb{E}_\theta(T))^2] = \mathrm{Var}_\theta(T),$$

ce qui conduit à la définition suivante.

Définition 8.2.3. *Soit T un estimateur de $f(\theta)$. On dit que T est un estimateur sans biais de variance minimum de $f(\theta)$ (en abrégé e.s.b.v.m.) si T est un e.s.b. de $f(\theta)$ et si, pour tout S e.s.b. de $f(\theta)$, on a, pour tout $\theta \in \Theta$, $Var_\theta(T) \leq Var_\theta(S)$.*

8.2.2. Exemple. Soit X un 1-échantillon de $B(n, \theta)$, $0 < \theta < 1$ inconnu. On veut estimer $f_1(\theta) = \theta$, $f_2(\theta) = \theta^2$, $f_3(\theta) = \theta - \theta^2$.

Notons d'abord que, si ϕ_1 et ϕ_2 sont deux e.s.b. de $f(\theta)$, on a, posant $\alpha = \phi_1 - \phi_2$, pour tout θ, $\mathbb{E}_\theta(\alpha(X)) = 0$, soit:

$$0 = \sum_{k=0}^{n} C_n^k \theta^k (1-\theta)^{n-k} \alpha(k) = (1-\theta)^n \sum_{k=0}^{n} C_n^k (\frac{\theta}{1-\theta})^k \alpha(k).$$

Donc, pour tout $u \in]0,1[$, $\sum_{k=0}^{n} C_n^k \alpha(k) u^k = 0$ et $\alpha \equiv 0$ i.e. $\phi_1 = \phi_2$. Un e.s.b. est donc unique et c'est un e.s.b.v.m.

(i) On sait que $\mathbb{E}_\theta(X) = n\theta$ d'où $\frac{X}{n}$ est un e.s.b. et donc un e.s.b.v.m. de θ.

(ii) On sait que $\mathrm{Var}_\theta(X) = n\theta(1-\theta)$ d'où $\mathbb{E}_\theta(X^2) = n^2\theta^2 + n\theta(1-\theta) = n(n-1)\theta^2 + n\theta$ et $\frac{X(X-1)}{n(n-1}$ est un e.s.b. et donc un e.s.b.v.m. de θ^2.

(iii) Il résulte de (i) et (ii) que $\mathbb{E}_\theta(\frac{X}{n} - \frac{X(X-1)}{n(n-1)}) = \theta - \theta^2$. Donc $\frac{X(n-X)}{n(n-1)}$ est un e.s.b. et aussi un e.s.b.v.m. de $\theta - \theta^2$.

8.2.3. Un critère général.

Proposition 8.2.4. *Soit T un e.s.b. de $f(\theta)$. C'est un e.s.b.v.m. ssi, pour toute statistique réelle U telle que, pour tout $\theta \in \Theta$, $\mathbb{E}_\theta(U) = 0$, on a, pour tout $\theta \in \Theta$, $\mathbb{E}_\theta(TU) = 0$.*

Preuve: (i) On suppose que T vérifie la condition ci-dessus. Soient S un e.s.b. de $f(\theta)$ et $U = S - T$. On a $\mathbb{E}_\theta(U) \equiv 0$ et

$$\text{Var}_\theta(S) = \text{Var}_\theta(T + U) = \text{Var}_\theta(T) + \text{Var}_\theta(U) + 2\text{Cov}_\theta(T, U) \geq \text{Var}_\theta(T)$$

puisque $\text{Cov}_\theta(T, U) = \mathbb{E}_\theta(TU) - \mathbb{E}_\theta(T)\mathbb{E}_\theta(U) \equiv 0$.

(ii) On suppose que T est un e.s.b.v.m. de $f(\theta)$. Soient U telle que $\mathbb{E}_\theta(U) \equiv 0$ et $S = T + \rho U$. Evidemment S est un e.s.b. de $f(\theta)$. On a, puisque $\mathbb{E}_\theta(U) \equiv 0$,

$$\text{Var}_\theta(S) = \text{Var}_\theta(T + \rho U) = \text{Var}_\theta(T) + 2\rho\mathbb{E}_\theta(TU) + \rho^2\text{Var}_\theta(U).$$

Supposons $\mathbb{E}_\theta(TU) > 0$. Choisissant $\rho < 0$ assez près de 0, on a $\text{Var}_\theta(S) < \text{Var}_\theta(T)$ ce qui contredit T e.s.b.v.m. On fait le même raisonnement si $\mathbb{E}_\theta(TU) < 0$ et finalement on obtient $\mathbb{E}_\theta(TU) \equiv 0$.

8.2.4. Applications.

(i) Soit $X_1, \ldots, X_n$ un n-échantillon de la loi de Poisson $\mathcal{P}(\theta)$, $\theta > 0$ inconnu. On veut estimer θ. La loi de $(X_1, \ldots, X_n)$ est

$$\mathbb{P}_\theta(X_1 = x_1, \ldots, X_n = x_n) = \mathrm{e}^{-n\theta}\frac{\theta^{x_1+\ldots+x_n}}{x_1! \ldots x_n!},\ x_k \in \mathbb{N}.$$

Puisque $\mathbb{E}_\theta(X_1) = \theta$, $\overline{X} = \frac{1}{n}\sum_{k=1}^n X_k$ est un e.s.b. de θ. Soit $U = U(x_1, \ldots, x_n)$, $x_k \in \mathbb{N}$, telle que $\mathbb{E}_\theta(U) \equiv 0$. On a alors, pour tout $\theta > 0$,

$$\sum_{x_1,\ldots,x_n} U(x_1, \ldots, x_n)\,\frac{\theta^{x_1+\ldots+x_n}}{x_1! \ldots x_n!} = 0. \tag{8.7}$$

Dérivant (8.7) en θ, on a, pour tout θ,

$$\sum_{x_1,\ldots,x_n} U(x_1, \ldots, x_n)(x_1 + \ldots + x_n)\,\frac{\theta^{x_1+\ldots+x_n}}{x_1! \ldots x_n!} = 0,$$

soit encore $\mathbb{E}_\theta(U\overline{X}) \equiv 0$. On applique la prop. 8.2.4 et $\overline{X}$ est un e.s.b.v.m. de θ.

(ii) Soit $X_1, \ldots, X_n$ un n-échantillon de la loi normale $N_1(m, \sigma^2)$, $\theta = (m, \sigma^2)$ inconnu. On veut estimer m et σ^2. On sait que la densité de $(X_1, \ldots, X_n)$ est

$$q_\theta(x_1, \ldots, x_n) = (2\pi\sigma^2)^{-n/2}\exp(-\frac{1}{2\sigma^2}\sum_{k=1}^n (x_k - m)^2).$$

Posant

$$\rho = \frac{1}{2\sigma^2}, \quad \overline{x} = \frac{1}{n}\sum_{k=1}^{n} x_k, \quad s_0^2 = \frac{1}{n-1}\sum_{k=1}^{n}(x_k - \overline{x})^2,$$

on a, puisque (lem.8.1.6) $\sum_{k=1}^{n}(x_k - m)^2 = \sum_{k=1}^{n}(x_k - \overline{x})^2 + n(\overline{x} - m)^2$,

$$q_\theta(x_1, \dots, x_n) = (\frac{\rho}{\pi})^{n/2} \exp(-\rho(n-1)s_0^2 - n\rho(\overline{x} - m)^2).$$

Soit $U = U(x_1, \dots, x_n)$ telle que $\mathbb{E}_\theta(U) \equiv 0$. Alors, pour tous m, ρ,

$$\int U(x_1, \dots, x_n) \exp(-\rho(n-1)s_0^2 - n\rho(\overline{x} - m)^2)\, dx_1 \dots dx_n = 0. \tag{8.8}$$

Dérivant (8.8) en m, on a, pour tous tous m, ρ,

$$\int U(x_1, \dots, x_n)(\overline{x} - m) \exp(-\rho(n-1)s_0^2 - n\rho(\overline{x} - m)^2)\, dx_1 \dots dx_n = 0. \tag{8.9}$$

Soit encore $\mathbb{E}_\theta(U(\overline{X} - m)) \equiv 0$ et, vu que $\mathbb{E}_\theta(U) \equiv 0$, $\mathbb{E}_\theta(U\overline{X}) \equiv 0$. Comme $\overline{X}$ est un e.s.b. de m, la prop. 8.2.4 implique que c'est un e.s.b.v.m.

Dérivant (8.9) en m, on a, pour tous tous m, ρ,

$$\int U(x_1, \dots, x_n)(1 + 2n\rho(\overline{x} - m)^2) \exp(-\rho(n-1)s_0^2 - n\rho(\overline{x} - m)^2)\, dx_1 \dots dx_n = 0,$$

d'où $\mathbb{E}_\theta((1 + 2n\rho(\overline{X} - m)^2)U) \equiv 0$ et $\mathbb{E}_\theta((\overline{X} - m)^2 U) \equiv 0$.

Dérivant (8.8) en ρ, on a, pour tous tous m, ρ,

$$\int U(x_1, \dots, x_n)((n-1)s_0^2 + n(\overline{x} - m)^2) \exp(-\rho(n-1)s_0^2 - n\rho(\overline{x} - m)^2)\, dx_1 \dots dx_n = 0$$

i.e. $\mathbb{E}_\theta(U((n-1)s^2 + n(\overline{X} - m)^2)) \equiv 0$ où $s^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{X})^2$. On a vu que $\mathbb{E}_\theta((\overline{X} - m)^2 U) \equiv 0$, on a donc $\mathbb{E}_\theta(Us^2) \equiv 0$. On sait (8.1.2) que s^2 est un e.s.b. de σ^2, c'est donc un e.s.b.v.m. (prop.8.2.4).

8.2.5. Consistance. Soit $(\mathcal{X}, \mathcal{A}, (\mathbb{P}_\theta)_{\theta\in\Theta})$ un modèle statistique.

Définition 8.2.5. *Une suite T_n d'estimateurs de $f(\theta)$ est dite consistante si, pour tout $\theta \in \Theta$, $T_n \to_n f(\theta)$ $\mathbb{P}_\theta$ p.s.*

Il est clair que cette définition a un sens si f est à valeurs $\mathbb{R}^p$ et alors T_n est une suite d'applications de $\mathcal{X}$ dans $\mathbb{R}^p$. Elle est surtout utile pour un modèle statistique associé (voir la def. 8.1.8) à un échantillon de taille infinie $X_1, \dots, X_n, \dots$ d'une loi μ_θ et des estimateurs T_n de la forme $T_n = \phi_n(X_1, \dots, X_n)$. Par exemple, si μ est une loi sur $\mathbb{R}$ admettant un moment d'ordre 2, $\overline{X}_n$ et s_n sont des estimateurs consistants de la moyenne et la variance de μ.

8.2.6. Méthode des moments. Soient $(\mu_\theta,\ \theta \in \Theta)$ une famille de probabilités sur $\mathbb{R}^d$, $(\mathcal{X}, \mathcal{A}, (\mathbb{P}_\theta)_{\theta\in\Theta})$ le modèle statistique associé à un échantillon de taille infinie de μ_θ (def. 8.1.8) et $f : \Theta \to \mathbb{R}^p$. On veut estimer $f(\theta)$. On considère des fonctions $g_1, \dots g_r$ de $\mathbb{R}^d$ dans $\mathbb{R}$ telles que, pour tout $\theta \in \Theta$ et pour $i = 1, \dots, r$, $\mathbb{E}_\theta(|g_i(X_1)|) < +\infty$ et on pose $m_i(\theta) = \mathbb{E}_\theta(g_i(X_1))$. On suppose que $f(\theta)$ peut s'écrire $f(\theta) = \phi(m_1(\theta), \dots, m_r(\theta))$ avec ϕ continue.

D'après la loi forte des grands nombres,

$$\text{pour tout } \theta \in \Theta, \text{ pour } i = 1, \dots, r,\ \ \hat{m}_i^n = \frac{1}{n}\sum_{k=1}^n g_i(X_k) \to_n m_i(\theta)\ \ \mathbb{P}_\theta \text{ p.s..}$$

Donc, si on pose,

$$T_n = \phi(\hat{m}_1^n, \dots, \hat{m}_r^n), \tag{8.10}$$

pour tout $\theta \in \Theta$, $T_n \to_n f(\theta)$, $\mathbb{P}_\theta$ p.s. i.e. T_n est une suite consistante d'estimateurs de $f(\theta)$. Donc, si n est asez grand, on peut utiliser T_n comme estimateur de $f(\theta)$.

Si $d = 1$, on peut choisir $g_1(u) = u, g_2(u) = u^2 \dots, g_r(u) = u^r$ et l'on a $m_i(\theta) = \mathbb{E}_\theta(X_1^r)$ d'où le nom de méthode des moments.

Exemple 1. Soit $X_1, \dots, X_n$ un n-échantillon de la loi sur $\mathbb{R}^+$ $G(a,c)$, $\theta = (a,c)$ inconnu. On a (voir 4.3.1.d):

$$m_1(\theta) = \mathbb{E}_\theta(X_1) = \frac{a}{c},\ m_2(\theta) = \mathbb{E}_\theta(X_1^2),\ \sigma^2(\theta) = \mathrm{Var}_\theta(X_1) = m_2(\theta) - (m_1(\theta))^2 = \frac{a}{c^2}.$$

Donc

$$a = \frac{(m_1(\theta))^2}{\sigma^2(\theta)},\ \ c = \frac{m_1(\theta)}{\sigma^2(\theta)}.$$

On a $\hat{m}_1 = \frac{1}{n}\sum_{k=1}^n X_k = \overline{X}$, $\hat{m}_2 = \frac{1}{n}\sum_{k=1}^n X_k^2$ et, posant

$$\hat{\sigma}^2 = \hat{m}_2 - (\hat{m}_1)^2 = \frac{1}{n}\sum_{k=1}^n X_k^2 - (\overline{X})^2 = \frac{1}{n}\sum_{k=1}^n (X_k - \overline{X})^2,$$

on obtient comme estimateurs de a et c:

$$\hat{a} = \frac{(\overline{X})^2}{\hat{\sigma}^2},\ \ \ \hat{c} = \frac{\overline{X}}{\hat{\sigma}^2}.$$

Exemple 2. Soit $X_1, \dots, X_n$ un n-échantillon de la loi sur $\mathbb{R}$ de densité q_θ donnée par

$$q_\theta(x) = \theta\, q_1(x) + (1-\theta)\, q_2(x),$$

où q_1 et q_2 sont des densités connues et $\theta \in [0,1]$ un paramètre inconnu qu'on veut estimer. Soit $(\Delta_i,\ i = 1, \dots, r)$ une partition de $\mathbb{R}$ en intervalles. On pose

$$\mu_{i,1} = \int_{\Delta_i} q_1(u)\, du,\ \ \mu_{i,2} = \int_{\Delta_i} q_2(u)\, du$$

et on suppose $\mu_{i,1} \neq \mu_{i,2}$ pour tout i. On choisit

$$g_i(u) = 1_{\{u \in \Delta_i\}}$$

et on a

$$m_i(\theta) = \mathbb{P}_\theta(X_1 \in \Delta_i) = \theta\mu_{i,1} + (1-\theta)\mu_{i,2}.$$

Il y a de multiple façon d'exprimer θ comme fonction des $m_i(\theta)$ puisque, pour chaque i, $\theta = \frac{m_i(\theta)-\mu_{i,2}}{\mu_{i,1}-\mu_{i,2}}$. On choisit

$$\theta = \frac{1}{r}\sum_{k=1}^{r} \frac{m_i(\theta) - \mu_{i,2}}{\mu_{i,1} - \mu_{i,2}}.$$

On obtient alors comme estimateur de θ:

$$\hat{\theta} = \frac{1}{r}\sum_{k=1}^{r} \frac{\hat{m}_i - \mu_{i,2}}{\mu_{i,1} - \mu_{i,2}}, \quad \hat{m}_i = \frac{1}{n}\sum_{k=1}^{n} 1_{\{X_k \in \Delta_i\}}.$$

8.2.7. Méthode du maximum de vraisemblance. Considérons le modèle statistique suivant. $\mathcal{X} = \{x_1, x_2\}$, $\Theta = \{\theta_1, \theta_2\}$,

$$\mathbb{P}_{\theta_1}(x_1) = \frac{1}{100},\ \mathbb{P}_{\theta_1}(x_2) = \frac{99}{100},\ \mathbb{P}_{\theta_2}(x_1) = \frac{99}{100},\ \mathbb{P}_{\theta_2}(x_2) = \frac{1}{100}.$$

On tire un point de $\mathcal{X}$ selon $\mathbb{P}_{\theta_i}$, $i = 1, 2$, inconnu. Supposons qu'on obtienne x_1. Il est naturel d'estimer θ par θ_2. Qu'a-t-on fait ? On a comparé $\mathbb{P}_{\theta_1}(x_1) = \frac{1}{100}$ et $\mathbb{P}_{\theta_2}(x_1) = \frac{99}{100}$ et on a choisi la valeur de θ rendant maximum la fonction $\theta \mapsto \mathbb{P}_\theta(x_1)$. C'est le principe de la méthode du maximum de vraisemblance.

Soit $(\mathcal{X}, \mathcal{A}, (\mathbb{P}_\theta)_{\theta\in\Theta})$ un modèle statistique. On suppose qu'il existe une mesure σ-finie μ sur $(\mathcal{X}, \mathcal{A})$ telle que, pour tout θ, $\mathbb{P}_\theta = f_\theta.\mu$ et on pose

$$L(x;\theta) = f_\theta(x). \tag{8.11}$$

La fonction $\theta \mapsto L(x;\theta)$ s'appelle la fonction de vraisemblance associée à x.

Définition 8.2.6. *Soit $T : \mathcal{X} \to \Theta$. On dit que T est un estimateur du maximum de vraisemblance de θ (en abrégé e.m.v.) si,*

$$\textit{pour tout } x \in \mathcal{X},\ \ L(x;T(x)) = \sup_{\theta\in\Theta} L(x;\theta). \tag{8.12}$$

Pour calculer un e.m.v., on est donc amené à chercher, pour tout $x \in \mathcal{X}$, pour quelle(s) valeur(s), $\theta \mapsto L(x;\theta)$ ou, ce qui revient au même, $\theta \mapsto \log L(x;\theta)$ est maximum. Si Θ est un ouvert de $\mathbb{R}^d$, si $L(x;\theta) \to 0$ lorsque θ tend vers le bord de Θ et si L est dérivable en θ, ces valeurs sont à chercher parmi les solutions de

$$\frac{\partial}{\partial\theta_i} \log L(x;\theta) = 0,\ \ i = 1, \ldots, d. \tag{8.13}$$

L'équation (8.13) s'appelle l'équation de vraisemblance.

Pour un échantillon de taille finie, il est difficile de justifier cette méthode. Par contre, pour un échantillon de taille infinie $X_1, \ldots, X_n, \ldots$ et sous des hypothèses relativement générales, il existe une suite T_n consistante (voir 8.2.5) d'estimateurs de θ, T_n étant un e.m.v. associé au n-échantillon $(X_1, \ldots, X_n)$.

8.2.8. Exemples.

(i) Soit $X_1, \ldots, X_n$ un n-échantillon de la loi sur $\mathbb{R}^+$ de densité $\theta \mathrm{e}^{-\theta x}$, $\theta > 0$ inconnu. Prenant $\mu = \lambda_+^{\otimes n}$, λ_+ mesure de Lebesgue sur $\mathbb{R}^+$, on a

$$L(x;\theta) = L(x_1, \ldots, x_n; \theta) = \theta^n \mathrm{e}^{-\theta(x_1+\ldots+x_n)}$$

et, posant $\overline{x} = \frac{1}{n}(x_1 + \ldots + x_n)$,

$$\log L(x;\theta) = n \log \theta - \theta n \overline{x}.$$

Alors $\frac{d}{d\theta} \log L(x;\theta) = \frac{n}{\theta} - n\overline{x} = 0$ pour $\theta = \hat{\theta} = 1/\overline{x}$. Vu que $L(x;\theta) \to 0$ lorsque $\theta \to 0$ et $\theta \to +\infty$, cette valeur correspond à un maximum est $1/\overline{x}$ est l'e.m.v. de θ.

(ii) Soit $X_1, \ldots, X_n$ un n-échantillon de $N_1(m, \sigma^2)$, $\theta = (m, \sigma^2)$ inconnu. On a

$$\log L(x;\theta) = \log L(x_1, \ldots, x_n; \theta) = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{k=1}^{n} (x_k - m)^2.$$

On en déduit (on considère σ^2 comme une variable)

$$\begin{aligned} \frac{\partial}{\partial m} \log L(x;\theta) &= \frac{1}{\sigma^2} \sum_{k=1}^{n} (x_k - m) \\ \frac{\partial}{\partial \sigma^2} \log L(x;\theta) &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{k=1}^{n} (x_k - m)^2. \end{aligned}$$

Alors $\frac{\partial}{\partial m} \log L(x;\theta) = \frac{\partial}{\partial \sigma^2} \log L(x;\theta) = 0$ a pour solution

$$\hat{m} = \frac{1}{n} \sum_{k=1}^{n} x_k = \overline{x}, \;\; \hat{\sigma}^2 = \frac{1}{n} \sum_{k=1}^{n} (x_k - \hat{m})^2 = \frac{1}{n} \sum_{k=1}^{n} (x_k - \overline{x})^2.$$

On vérifie que ces valeurs correspondent bien à un maximum. L'e.m.v. de (m, σ^2) est donc $(\overline{X}, \hat{s}^2)$ où $\hat{s}^2 = \frac{1}{n} \sum_{k=1}^{n} (X_k - \overline{X})^2$. Noter que $\hat{s}^2 = \frac{n-1}{n} s^2$ n'est pas sans biais.

8.3. Intervalle de confiance

On considère un modèle statistique $(\mathcal{X}, \mathcal{A}, (\mathbb{P}_\theta)_{\theta \in \Theta})$ et une application mesurable f de Θ dans $\mathbb{R}$. Plutôt que d'estimer $f(\theta)$ par un nombre $T(x)$ qui est probablement voisin de $f(\theta)$ mais pratiquement jamais égal à $f(\theta)$, on peut envisager de répondre

$f(\theta) \in I(x)$, $I(x)$ étant un intervalle dépendant du point tiré x et de préciser cette réponse en disant que $f(\theta) \in I(x)$ avec une probabilité au moins égale à $0,9$ ou $0,95 \ldots \ldots$.

8.3.1. Ceci conduit à:

Définition 8.3.1. *On appelle intervalle de confiance de niveau* $1-\alpha$ *pour* $f(\theta)$ *une famille d'intervalles* $(I(x),\, x \in \mathcal{X})$ *telles que, pour tout* $\theta \in \Theta$,

$$\mathbb{P}_\theta(f(\theta) \in I(X)) \geq 1-\alpha.$$

Evidemment une deuxième notion intervient pour juger de la qualité d'un intervalle de confiance, à savoir sa longueur et, plus on voudra α petit, plus l'intervalle sera long.

8.3.2. Fonction pivotale. On présente un procédé relativement général pour construire des intervalles de confiance. On appellera fonction pivotale monotone une application mesurable $g(x,u)$ de $\mathcal{X} \times \mathbb{R}$ dans $\mathbb{R}$ telle que
(i) pour tout $\theta \in \Theta$, la v.a. $g(X, f(\theta))$ suit une loi μ indépendante de θ,
(ii) pour tout $x \in \mathcal{X}$, $u \mapsto g(x,u)$ est strictement monotone.

On choisit alors $a < b$ tels que $\mu(]a,b[) = 1-\alpha$, on a donc, pour tout $\theta \in \Theta$, $\mathbb{P}_\theta(g(X,f(\theta)) \in]a,b[) = \mu(]a,b[) = 1-\alpha$. Mais, vu la monotonie, $\{g(X,f(\theta)) \in]a,b[\} = \{f(\theta) \in]A(X),B(X)[\}$ et $I(x) =]A(x),B(x)[$ est un intervalle de confiance de niveau $1-\alpha$ pour $f(\theta)$.

Exemple. Soit $X_1,\ldots,X_n$ un n-échantillon de $N_1(\theta,\sigma^2)$, σ^2 étant connu et θ inconnu. Alors $\overline{X} \sim N_1(\theta, \frac{\sigma^2}{n})$ et

$$\sqrt{n}\,\frac{\overline{X}-\theta}{\sigma} \sim N_1(0,1).$$

Donc $g(x,\theta) = \sqrt{n}\,\frac{\overline{x}-\theta}{\sigma}$ est une fonction pivotale monotone.

Etant donné α, on choisit $c = c(\alpha)$ dans une table de loi normale telle que $\frac{1}{\sqrt{2\pi}}\int_{-c}^{c} \mathrm{e}^{-t^2/2}\,dt = 1-\alpha$ et on a , pour tout $\theta \in \mathbb{R}$,

$$\mathbb{P}_\theta(\sqrt{n}\,\frac{|\overline{X}-\theta|}{\sigma} < c) = \mathbb{P}_\theta(\theta \in]\overline{X} - \frac{c\sigma}{\sqrt{n}}, \overline{X} + \frac{c\sigma}{\sqrt{n}}[) = 1-\alpha.$$

Evidemment, dans la plupart des cas, σ^2 n'est pas connu. On peut envisager de remplacer σ par son estimé s ce qui conduit à étudier la distribution de $\sqrt{n}\,\frac{\overline{X}-\theta}{s}$.

8.3.3. Echantillons gaussiens.

Définition 8.3.2. *Soit* $X_1,\ldots,X_n$ *un* n*-échantillon de* $N_1(0,1)$. *On appelle loi de chi-carré à* n *degrés de liberté et on note* χ^2_n *la loi de* $X_1^2 + \ldots + X_n^2$.

On sait (4.6.1) que $X_1^2 \sim G(\frac{1}{2},\frac{1}{2})$ donc (5.2.2.d) $X_1^2 + \ldots + X_n^2 \sim G(\frac{n}{2},\frac{1}{2})$ et la densité de la loi χ^2_n est:

$$\phi(x) = \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}\mathrm{e}^{-\frac{x}{2}}x^{\frac{n}{2}-1}1_{\mathbb{R}^+}(x). \tag{8.14}$$

Définition 8.3.3. *Soient X et Y deux v.a.r. indépendantes avec $X \sim N_1(0,1)$ et $Y \sim \chi^2_n$. On appelle loi de Student à n degrés de liberté et on note t_n la loi de*

$$T = \frac{X}{\sqrt{Y/n}}.$$

Un calcule facile montre que la loi t_n a pour densité:

$$h(t) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n+1}\,\Gamma(\frac{n}{2})}(1+\frac{t^2}{n})^{-\frac{n+1}{2}} \tag{8.15}$$

Théorème 8.3.4. *Soit $X_1,\dots,X_n$ un n-échantillon de $N_1(m,\sigma^2)$. Alors $\overline{X}$ et s^2 définis par (8.4) et (8.5) sont indépendants, $\overline{X} \sim N_1(m,\frac{\sigma^2}{n})$ et $(n-1)\frac{s^2}{\sigma^2} \sim \chi^2_{n-1}$. En particulier $\sqrt{n}\,\frac{\overline{X}-m}{s} \sim t_{n-1}$.*

Preuve: A. On suppose $m=0$ et $\sigma^2=1$. Alors $X=(X_1,\dots,X_n) \sim N_n(0,I_n)$. Soient A une matrice orthogonale $n\times n$ de la forme

$$A = \begin{pmatrix} \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \end{pmatrix}$$

et $Y=(Y_1,\dots,Y_n)=AX$. On a $Y \sim N_n(0,I_n)$ puisque $K(Y)=A\,K(X)\,A^{\mathsf{T}} = A\,A^{\mathsf{T}} = I_n$, $Y_n = \frac{1}{\sqrt{n}}(X_1+\ldots+X_n) = \sqrt{n}\,\overline{X}$ et, vu que $||X||^2 = ||AX||^2 = ||Y||^2$,

$$(n-1)s^2 = \sum_{k=1}^{n}(X_k-\overline{X})^2 = \sum_{k=1}^{n} X_k^2 - n(\overline{X})^2 = \sum_{k=1}^{n} Y_k^2 - Y_n^2 = \sum_{k=1}^{n-1} Y_k^2.$$

Ceci implique que $\overline{X} = \frac{1}{\sqrt{n}}Y_n \sim N_1(0,\frac{1}{\sqrt{n}})$ et est indépendant de $(n-1)s^2 = \sum_{k=1}^{n-1} Y_k^2$ qui suit χ^2_{n-1}.

B. On revient au cas général. On pose $Z_k = \sigma^{-1}(X_k-m)$. Alors $Z=(Z_1,\dots,Z_n)$ un n-échantillon de $N_1(0,1)$, $\overline{X} = m+\sigma\overline{Z}$ et

$$(n-1)s_X^2 = \sum_{k=1}^{n}(X_k-\overline{X})^2 = \sigma^2\sum_{k=1}^{n}(Z_k-\overline{Z})^2 = \sigma^2(n-1)s_Z^2.$$

D'où $\sqrt{n}\,\frac{\overline{X}-m}{\sigma} \sim N_1(0,1)$, $(n-1)\frac{s^2}{\sigma^2} \sim \chi^2_{n-1}$ et sont indépendants. Appliquant la def. 8.3.3, on obtient la dernière affirmation. $\diamond$

Application. Soit $X_1,\dots,X_n$ un n-échantillon de $N_1(m,\sigma^2)$, $\theta=(m,\sigma^2)$ inconnu. On cherche des intervalles de confiance pour m et σ^2.

(i) On choisit $c = c(\alpha)$ tel que $\mathbb{P}(|T| < c) = 1 - \alpha$ où $T \sim t_{n-1}$. Alors (th.8.3.4), pour tout $\theta = (m, \sigma^2)$,

$$\mathbb{P}_\theta(\sqrt{n}\,|\frac{\overline{X} - m}{s}| < c) = \mathbb{P}_\theta(m \in]\,\overline{X} - \frac{cs}{\sqrt{n}}, \overline{X} + \frac{cs}{\sqrt{n}}\,[) = 1 - \alpha.$$

(ii) On choisit $a < b$ tels que $\mathbb{P}(a < Y < b) = 1 - \alpha$ où $Y \sim \chi^2_{n-1}$. Alors (th.8.3.4), pour tout $\theta = (m, \sigma^2)$,

$$\mathbb{P}_\theta(a < (n-1)\frac{s^2}{\sigma^2} < b) = \mathbb{P}_\theta(\sigma^2 \in]\,\frac{(n-1)s^2}{b}, \frac{(n-1)s^2}{a}\,[) = 1 - \alpha.$$

8.3.4. Intervalle de confiance asymptotique. Un intervalle de confiance asymptotique de niveau $1 - \alpha$ pour $f(\theta)$ est une suite de familles d'intervalles $(I_n(x),\ x \in \mathcal{X})$ telle que, pour tout θ,

$$\mathbb{P}_\theta(f(\theta) \in I_n(X)\,) \to_n 1 - \alpha.$$

Pour construire de tels intervalles, on peut utiliser (rappelons que $\overline{X}_n$ et s_n ont été définis en (8.4) et (8.5)):

Proposition 8.3.5. *Soit* $(X_n,\ n \geq 1)$ *une suite de v.a.r. de carré intégrable indépendantes et de même loi. On pose* $m = \mathbb{E}(X_1)$, $\sigma^2 = Var(X_1)$ *qu'on suppose* > 0. *Alors* $\sqrt{n}\,\frac{\overline{X}_n - m}{s_n} 1_{\{s_n > 0\}}$ *converge en loi vers* $N_1(0,1)$.

Preuve: On a

$$\sqrt{n}\,\frac{\overline{X}_n - m}{s_n} 1_{\{s_n > 0\}} = \sqrt{n}\,\frac{\overline{X}_n - m}{\sigma}.\frac{\sigma}{s_n} 1_{\{s_n > 0\}}.$$

D'une part $\sqrt{n}\,\frac{\overline{X}_n - m}{\sigma}$ converge en loi vers $N_1(0,1)$ (th. 7.3.1). D'autre part $s_n \to_n \sigma$ p.s. (8.1.3) et donc $\frac{\sigma}{s_n} 1_{\{s_n > 0\}} \to_n 1$ p.s. On conclut par la prop. 7.2.7. $\diamond$

Soit $(X_n,\ n \geq 0)$ un échantillon de taille infinie d'une loi μ sur $\mathbb{R}$ de densité q de moyenne m avec $\int x^2\, d\mu(x) < +\infty$. On a alors $\mathbb{P}(X_1 = X_2) = 0$ et, a fortiori, $P(s_n > 0) = 1$. On choisit $c = c(\alpha)$ tel que $(2\pi)^{-1/2} \int_{-c}^{c} \mathrm{e}^{-t^2/2}\, dt = 1 - \alpha$. Donc, vu les prop. 8.3.5 et 7.2.4,

$$\mathbb{P}(\sqrt{n}\,\frac{|\overline{X}_n - m|}{s_n} < c) = \mathbb{P}(\,m \in]\overline{X}_n - \frac{cs_n}{\sqrt{n}}, \overline{X}_n + \frac{cs_n}{\sqrt{n}}[\,) \to_n 1 - \alpha.$$

On a construit un intervalle de confiance asymptotique de niveau $1 - \alpha$ pour m.

8.4. Tests

8.4.1. Généralités. Soit $(\mathcal{X}, \mathcal{A}, (\mathbb{P}_\theta)_{\theta \in \Theta})$ un modèle statistique. On suppose que $\Theta = H_0 \cup H_1$ avec $H_0 \cap H_1 = \emptyset$. Il s'agit, à la vue du point x tiré selon $\mathbb{P}_\theta$, θ inconnu, de décider si $\theta \in H_0$ ou non. Cela s'appelle tester l'hypothèse H_0 contre l'hypothèse H_1. Un test de H_0 contre H_1 est donc un sous-ensemble W de $\mathcal{X}$, appelé région critique

ou région de rejet. Si le point tiré x appartient à W, on rejette l'hypothèse H_0, si $x \notin W$, on l'accepte.

Il y a deux types d'erreur.

(i) Si $\theta \in H_0$, $\mathbb{P}_\theta(W)$ représente la probabilité de rejeter à tort H_0, c'est l'erreur de première espèce.

(ii) Si $\theta \in H_1$, $\mathbb{P}_\theta(W^c) = 1 - \mathbb{P}_\theta(W)$ représente la probabilité d'accepter à tort H_0, c'est l'erreur de deuxième espèce.

Dans la théorie classique des tests, on fixe un seuil maximum à l'erreur de première espèce à savoir $0,1,\ 0,05,\ 0,01\ldots$ ce qui conduit à la définition:

Définition 8.4.1. *Soit W la région critique d'un test de H_0 contre H_1. La quantité*

$$\alpha = \alpha(W) = \sup_{\theta \in H_0} \mathbb{P}_\theta(W) \tag{8.16}$$

s'appelle le niveau du test. La fonction de H_1 dans $[0,1]$, $\theta \mapsto \mathbb{P}_\theta(W)$, s'appelle la fonction puissance du test.

Le niveau étant fixé, il s'agit de trouver des régions W telles que, pour $\theta \in H_1$, $\mathbb{P}_\theta(W)$ soit le plus grand possible. Comme en estimation, il est quasiment impossible de trouver un test optimal si on ne restreint pas la classe considérée.

Définition 8.4.2. *Soit W la région critique d'un test de H_0 contre H_1. On dit que le test est sans biais au seuil α s'il est de niveau inférieur ou égal à α et si, pour tout $\theta \in H_1$, $\mathbb{P}_\theta(W) \geq \alpha$.*

Définition 8.4.3. *Un test de région critique W de niveau α de H_0 contre H_1 est dit uniformément plus puissant sans biais (en abrégé U.P.P.S.B.) s'il est sans biais au seuil α et si, pour tout test de région critique W' sans biais au seuil α de H_0 contre H_1, on a, pour tout $\theta \in H_1$, $\mathbb{P}_\theta(W) \geq \mathbb{P}_\theta(W')$.*

Terminons ces généralités par un mot de la théorie asymptotique.

Définition 8.4.4. *Une suite de tests de H_0 contre H_1 de région critique W_n est dite consistante de niveau asymptotique α si, pour tout $\theta \in H_0$, $\mathbb{P}_\theta(W_n) \to_n \alpha$ et si, pour tout $\theta \in H_1$, $\mathbb{P}_\theta(W_n) \to_n 1$.*

8.4.2. Le lemme de Neyman-Pearson. Dans le cas d'hypothèses simples i.e. réduites à un point, il est facile d'avoir un test optimal.

Lemme 8.4.5. *On suppose $\Theta = \{\theta_0, \theta_1\}$ et $\mathbb{P}_{\theta_0} = h_0.\mu$, $\mathbb{P}_{\theta_1} = h_1.\mu$. Alors $W = \{x,\ h_1(x) \geq \lambda h_0(x)\}$ est, pour tout $\lambda > 0$, la région critique de $\theta = \theta_0$ contre $\theta = \theta_1$ le plus puissant à son niveau.*

Preuve: Soit D la région critique d'un autre test tel que $\mathbb{P}_{\theta_0}(D) \leq \mathbb{P}_{\theta_0}(W)$. On remarque que $(1_W - 1_D)(h_1 - \lambda h_0) \geq 0$ d'où $\int (1_W - 1_D)(h_1 - \lambda h_0)\, d\mu \geq 0$ et

$$\mathbb{P}_{\theta_1}(W) - \mathbb{P}_{\theta_1}(D) = \int (1_W - 1_D) h_1\, d\mu \geq \lambda \int (1_W - 1_D) h_0\, d\mu = \lambda(\mathbb{P}_{\theta_0}(W) - \mathbb{P}_{\theta_0}(D)) \geq 0.$$

Le test de région critique W est plus puissant que le test de région critique D. $\diamond$

Pour utiliser le lem.8.4.5, étant donné α, on détermine λ par la condition

$$\mathbb{P}_{\theta_0}(\{h_1 \geq \lambda h_0\}) = \int_{\{h_1 \geq \lambda h_0\}} h_0\, d\mu = \alpha.$$

8.4.3. Tests sur échantillons gaussiens.

1. Soit $X_1, \ldots, X_n$ un n-échantillon de $N_1(m, \sigma^2)$ avec $\theta = (m, \sigma^2)$ inconnu. Soit $m_0 \in \mathbb{R}$ fixé. Il s'agit de tester $H_0 :\ m = m_0$ contre $H_1 :\ m \neq m_0$. On sait (def. 8.3.3) que $\sqrt{n}\, \frac{\overline{X} - m}{s} \sim t_{n-1}$. Considérons

$$W = \{ \sqrt{n}\, |\frac{\overline{X} - m_0}{s}| > c \}.$$

Sous H_0 i.e. si $m = m_0$, $\mathbb{P}_\theta(W) = \mathbb{P}(|T| > c)$ où $T \sim t_{n-1}$. On détermine $c = c(\alpha)$ comme solution de $\mathbb{P}(|T| > c) = \alpha$ à l'aide d'une table de la loi de Student et W est la région critique d'un test de niveau α de $m = m_0$ contre $m \neq m_0$. On peut montrer que ce test est U.P.P.S.B.

2. Soient $X_1, \ldots, X_n$ un n-échantillon de $N_1(m_1, \sigma^2)$ et $Y_1, \ldots, Y_r$ un r-échantillon de $N_1(m_2, \sigma^2)$. On suppose $(X_i,\ 1 \leq i \leq n)$ et $(Y_j,\ 1 \leq j \leq r)$ indépendants. On a $\theta = (m_1, m_2, \sigma^2)$ inconnu. Il s'agit de tester $H_0 :\ m_1 = m_2$ contre $H_1 :\ m_1 \neq m_2$. On pose

$$\overline{X} = \frac{1}{n} \sum_{i=1}^n X_i,\ \ s_1^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X})^2,\ \ \overline{Y} = \frac{1}{r} \sum_{j=1}^r Y_j,\ \ s_2^2 = \frac{1}{r-1} \sum_{j=1}^r (Y_j - \overline{Y})^2.$$

Lemme 8.4.6. *Sous les hypothèses ci-dessus, on a, si $m_1 = m_2$,*

$$Z = \sqrt{\frac{n+r-2}{\frac{1}{n} + \frac{1}{r}}}\ \frac{\overline{X} - \overline{Y}}{\sqrt{(n-1)s_1^2 + (r-1)s_2^2}} \sim t_{n+r-2}.$$

Preuve: D'une part $\overline{X} \sim N_1(m_1, \frac{\sigma^2}{n})$, $\overline{Y} \sim N_1(m_2, \frac{\sigma^2}{r})$ et, vu l'indépendance (prop. 5.2.7), $\overline{X} - \overline{Y} \sim N_1(m_1 - m_2, \frac{\sigma^2}{n} + \frac{\sigma^2}{r})$ et, si $m_1 = m_2$, $\frac{\overline{X} - \overline{Y}}{\sigma\sqrt{\frac{1}{n} + \frac{1}{r}}} \sim N_1(0, 1)$.

D'autre part $(n-1)\frac{s_1^2}{\sigma^2} \sim \chi^2_{n-1}$, $(r-1)\frac{s_2^2}{\sigma^2} \sim \chi^2_{r-1}$ et, vu l'indépendance, $(n-1)\frac{s_1^2}{\sigma^2} + (r-1)\frac{s_2^2}{\sigma^2} \sim \chi^2_{n+r-2}$.

Puisque $(\overline{X}, \overline{Y})$ est indépendant de (s_1^2, s_2^2), on peut appliquer la def. 8.3.3. $\diamond$

Posons $W = \{|Z| > c\}$ où $\mathbb{P}(|T| > c) = \alpha$, $T \sim t_{n+r-2}$. On a, sous H_0 i.e. si $m_1 = m_2$, $\mathbb{P}_\theta(W) = \mathbb{P}(|T| > c) = \alpha$ et W est la région critique d'un test de niveau α de $m_1 = m_2$ contre $m_1 \neq m_2$. On peut montrer que ce test est U.P.P.S.B.

Remarque. Le lecteur peut noter une grande ressemblance entre la construction de tests et celle d'intervalles de confiance. Cela n'a rien d'étonnant. En effet, étant donné un modèle stastique $(\mathcal{X}, \mathcal{A}, (\mathbb{P}_\theta)_{\theta\in\Theta})$, soit $(W_a,\, a \in \mathbb{R})$ une famille de sous ensembles de $\mathcal{X}$ (avec $W_a \in \mathcal{A}$ mais nous n'insistons pas sur ce point). On pose, pour $x \in \mathcal{X}$, $S(x) = \{a,\, x \notin W_a\}$. Evidemment $W_a = \{x,\, a \notin S(x)\}$ et, pour tout $\theta \in \Theta$ et tout $a \in \mathbb{R}$,

$$\mathbb{P}_\theta(W_a) = \mathbb{P}_\theta(x,\, a \notin S(x)) = 1 - \mathbb{P}_\theta(x,\, a \in S(x)). \tag{8.17}$$

Soit $f : \Theta \to \mathbb{R}$. Il résulte de (8.17) que si, pour tout a, W_a est la région critique d'un test de niveau α de $H_0 :\ f(\theta) = a$ contre $H_1 :\ f(\theta) \neq a$, alors $S(x) = \{a,\, x \notin W_a\}$ est une région de confiance de niveau $1-\alpha$ pour $f(\theta)$ (c'est la même définition que celle d'un intervalle de confiance mais a priori $S(x)$ n'est pas un intervalle). De même si $(S(x),\, x \in \mathcal{X})$ est une région de confiance de niveau $1-\alpha$ pour $f(\theta)$, $W_a = \{x,\, a \notin S(x)\}$ est la région critique d'un test de niveau α de $H_0 :\ f(\theta) = a$ contre $H_1 :\ f(\theta) \neq a$.

8.4.4. Test d'adéquation. Soient E un ensemble fini qu'on peut supposer être $\{1, \ldots, r\}$ et Π l'ensemble des probabilités sur E. On fixe $p \in \Pi$ telle que, pour tout j, $p_j > 0$. On considère un échantillon $X_1, \ldots, X_n, \ldots$ d'une loi $\pi \in \Pi$ inconnue et on veut tester $H_0 :\ \pi = p$ contre $H_1 :\ \pi \neq p$. Posant

$$N_n^j = \sum_{k=1}^n 1_{\{j\}}(X_k), \tag{8.18}$$

Pearson a proposé un test à partir des fréquences $\frac{1}{n}N_n^j$ d'observation des points j, $j = 1, \ldots, r$ qui repose sur:

Proposition 8.4.7. *Soit $(X_n,\, n \geq 1)$ une suite de v.a. indépendantes à valeurs E de même loi π. On pose*

$$T_n = \frac{1}{n}\sum_{j=1}^r \frac{(N_n^j - np_j)^2}{p_j} = \sum_{j=1}^r \frac{n}{p_j}(\frac{N_n^j}{n} - p_j)^2. \tag{8.19}$$

(i) Si $\pi = p$, T_n converge en loi vers χ^2_{r-1}.

(ii) Si $\pi \neq p$, T_n converge p.s. vers $+\infty$.

Preuve: (i) Supposons $\pi = p$. On a

$$T_n = |\frac{1}{\sqrt{n}}\sum_{k=1}^n U_k|^2,\ \ U_k = (\,\frac{1}{\sqrt{p}_1}(1_{\{1\}}(X_k) - p_1), \ldots, \frac{1}{\sqrt{p}_r}(1_{\{r\}}(X_k) - p_r)\,).$$

Les vecteurs aléatoires $U_1, \dots, U_n, \dots$ sont indépendants de même loi avec $\mathbb{E}(U_1) = 0$ et un calcul facile montre que

$$K(U_1) = I_r - aa^{\mathrm{T}}, \ \ a^{\mathrm{T}} = (\sqrt{p}_1 \ \dots \ \sqrt{p}_r).$$

Le th. 7.3.1 implique que $\frac{1}{\sqrt{n}}\sum_{k=1}^n U_k$ converge en loi vers $N_r(0, I_r - aa^{\mathrm{T}})$. Alors (prop. 7.2.3) $T_n = |\frac{1}{\sqrt{n}}U_n|^2$ converge en loi vers $|Y|^2$ où $Y \sim N_r(0, I_r - aa^{\mathrm{T}})$. Vu que $|a| = 1$, il existe une matrice A orthogonale $r \times r$ telle que $Aa = (0 \dots 01)^{\mathrm{T}}$ et posons $Z = AY$. On a

$$K(Z) = AK(Y)A^{\mathrm{T}} = I_r - (Aa)(Aa)^{\mathrm{T}} = \begin{pmatrix} I_{r-1} & 0 \\ 0 & 0 \end{pmatrix}$$

et $|Y|^2 = |Z|^2 \sim \chi^2_{r-1}$.

(ii) Supposons $\pi \neq p$. D'après la loi des grands nombres, $\frac{N_n^j}{n} - p_j \to_n \pi_j - p_j$ qui est $\neq 0$ pour au moins un j et $T_n \to_n +\infty$ p.s. $\diamond$

Considérons maintenant la région critique $W_n = \{T_n \geq c\}$ où $c = c(\alpha)$ est tel que $\mathbb{P}(X \geq c) = \alpha$, $X \sim \chi^2_{r-1}$. On a, vu les prop. 8.4.7 et 7.2.4, $\mathbb{P}_p(W_n) \to_n \alpha$ et, pour $\pi \neq p$, $\mathbb{P}_\pi(W_n) \to_n 1$. On a construit un test consistant de niveau asymptotique α (def. 8.4.4) de $H_0 : \pi = p$ contre $H_1 ; \pi \neq p$.

Ce test est susceptible de nombreuses généralisations pour lesquelles nous renvoyons aux ouvrages spécialisés. Par exemple, soit $X_1, \dots, X_n$ un échantillon d'une loi μ inconnue sur $(E, \mathcal{E})$. On veut tester $\mu = \mu_0$ contre $\mu \neq \mu_0$, μ_0 probabilité donnée. On peut partager E en r ensembles disjoints $E_1, \dots, E_r$ d'union E (on a intérêt à choisir $\mu_0(E_j)$ voisin de $\frac{1}{r}$) et tester à l'aide du test précédent $H_0 : \ \mu(E_j) = \mu_0(E_j)$ pour $j = 1, \dots, r$ contre $H_1 : \ \mu(E_j) \neq \mu_0(E_j)$ pour au moins un j.

Annexe A

Index des notations

1.2.3 renvoie chapitre 1, section 2, sous-section 3.

Annexe B

Index des termes

1.2.3 renvoie chapitre 1, section 2, sous-section 3.